DOCUMENT RESUME

ED 205 715 CE 029 684

AUTHOR	Chalupsky, Albert B.; And Others
TITLE	Competency Measurement in Vocational Education: A Review of the State of the Art.
INSTITUTION	American Institutes for Research, Washington, D.C.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, D.C.
PUB DATE	Jun 81
CONTRACT	300-79-0735
NOTE	77p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Competence; *Competency Based Education; Criterion Referenced Tests; *Minimum Competency Testing; Performance; Performance Tests; Student Evaluation; Test Construction; *Vocational Education
IDENTIFIERS	*Competency Tests

ABSTRACT

Intended to assist those working in the field of occupational competency testing, this report summarizes major milestones in the history of competency measurement in vocational education and overviews the current state of the art. To provide an orientation to the environment that stimulated the current interest in occupational competency testing, chapters 2 and 3 review competency-based vocational education and discuss evaluation concerns in vocational education. Next, chapter 4 presents some background information along with definitions of a few key terms used in competency testing. The current efforts of a number of organizations in the field of competency testing are reviewed in chapter 5. They are organized into three categories: Department of Education-funded organizations, state agencies and consortia, and job performance assessment in the military. Chapter 6 describes recent developments that merit attention from test developers in vocational education--simulations, adaptive paper-and-pencil tests, confidence testing, and Rasch modeling. The technical and legal problems in setting standards of performance on competency tests are discussed in chapter 7. A 10-page list of references is provided. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED205715

COMPETENCY MEASUREMENT IN VOCATIONAL EDUCATION:

A REVIEW OF THE STATE OF THE ART

Albert B. Chalupsky
Linda Phillips-Jones
Malcolm N. Danoff

JUNE 1981

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EG029684

AMERICAN INSTITUTES FOR RESEARCH

Prepared for:
Office of Vocational and Adult Education
U.S. Department of Education
Washington, D.C. 20202

Prepared under Contract Number 300-79-0735

DEVELOPMENT OF COMPETENCY MEASURES FOR VOCATIONAL SKILL AREAS

The research reported herein was performed pursuant to Contract No. 300-79-0735 with the U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Department of Education position or policy.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
	Why This Report?	1
	What This Report Is Not	1
II.	COMPETENCY-BASED VOCATIONAL EDUCATION	4
	Definition	4
	Distinguishing Features of CBVE Programs	5
	Effects of Competency-Based Vocational Education	6
	Selecting Competencies for CBVE	9
	Minimums Versus Maximums	12
	Implementing a CBVE Program	12
III.	EVALUATION CONCERNS IN VOCATIONAL EDUCATION	14
	Continued Criticism of Vocational Education Evaluation	14
	Vocational Education Evaluation Begins to Mature	16
	Increased Emphasis on Competency-Based Learner Outcomes	17
IV.	BACKGROUND OF OCCUPATIONAL COMPETENCY TESTING	19
	A Bit of History	19
	Some Definitions	21
V.	CURRENT ACTIVITIES IN OCCUPATIONAL COMPETENCY MEASUREMENT	25
	Department of Education-Funded Organizations	25
	State Agencies and Consortia	28
	Job Performance Assessment in the Military	36
VI.	SOME RECENT DEVELOPMENTS IN OCCUPATIONAL COMPETENCY TESTING	45
	Simulations	45
	Adaptive Paper-and-Pencil Tests	47
	Confidence Testing	49
	The Rasch Model and Latent Trait Theory	50
VII.	SETTING PROFICIENCY STANDARDS ON COMPETENCY TESTS	53
	Technical Problems in Setting Performance Standards	56
	Legal Considerations in Setting Test-Score Standards	60
	Epilogue	61
REFERENCES		62

LIST OF TABLES

Table No.		Page
1	Occupations Covered by NOCTI Teacher Occupational Competency Tests	30
2	Student Occupational Competency Achievement Tests Under Development by the NOCTI Consortium	31
3	The Ohio Vocational Achievement Tests	33
4	Florida Occupational Achievement Tests Under Development	35

ACKNOWLEDGMENTS

I would like to express my appreciation to the many individuals who contributed to this report.

To Linda Phillips-Jones who authored Chapters II and III and to Malcolm Danoff who authored the section of Chapter V covering military research and development in the field of competency testing.

To Brent Baxter, William Clemans, Robert Krug, and Marion Shaycoft of AIR and to Steven Zwillinger, the Department of Education Project Officer, for their review of the manuscript and their many helpful suggestions.

To John French and Audrey Johnston of AIR's Technical Library, for the assistance provided in obtaining many of the references.

To Ruth Lewis and Jean Reynolds who patiently set up the various stages of word processing the manuscript, through several revisions to final storage of the report; and to Marie Petrano and John Porter who assisted in the editorial review of the report.

Finally, to Carrie Davis who directed the typing and reproduction of the various versions of this report, kept track of the hundreds of references that were considered and those that finally were cited in the text, and who read each version of the report and offered many helpful suggestions.

The quality of this report is due in large measure to the generous help of the above individuals. Needless to say, any errors there may be in the report are the responsibility of the authors.

Albert B. Chalupsky, Director
Competency Measures for Vocational
Skill Areas Project

I. INTRODUCTION

Why This Report?

With the current emphasis on competency-based vocational education, and the evaluation of its effectiveness, has come a renewed interest in the objective measurement of occupational competencies. Evidence of this interest can be seen in the growing number of test development efforts underway today in the area of vocational education. This report summarizes the major milestones in the history of competency measurement in vocational education and presents an overview of the current state of the art.

The resource review on which this report is based was conducted as part of a three-year project supported by the Bureau of Occupational and Adult Education of the U.S. Office of Education (now the Office of Vocational and Adult Education of the Department of Education). The objectives of this project are to develop, field test, validate, and disseminate competency tests in 14 selected occupational areas and to design and help implement a program for continuing competency test development on a self-supporting basis.[1] While the major purpose of this review of previous research and development was to ensure that project staff capitalized on the latest experiences in developing and evaluating occupational competency measures, the highlights of this literature search are presented here to assist others who are working, or are planning to work, in the field of competency measures for vocational education. Specifically, the intent is to provide a review of occupational competency testing, including a summary of some of the major efforts under way today and some of the methodological developments that should be of interest to those working in this area.

What This Report Is Not

This report is not intended to serve as a basic reference guide or handbook on competency testing, nor does it pretend to encompass all of the issues important to testing in general. While the field is not particularly rich in background guides, there are nevertheless a number of useful references. For those interested in a comprehensive review of theory and

---

[1] Additional information on this project is presented in Chapter V.

technique in educational testing generally, two classic references are recommended: Educational Measurement, the first edition edited by Lindquist (1951) and the second edition, by Thorndike (1971), both sponsored by the American Council of Education. Another reference of particular value is the Eighth Mental Measurements Yearbook (Buros, 1978). Reviews and evaluations of a sizable number of vocational tests are presented in Volume II of the yearbook. For those interested in test development handbooks directed specifically to vocational education, Boyd and Shimberg (1971), Panitz and Olivo (1971), and Erickson and Wentling (1976) are suggested. Finally, readers who are particularly concerned about criterion-referenced testing should consult Shaycoft's definitive handbook (1979).

This report is not intended to overlap the above references nor does it focus on the problems of occupational licensing or certification, per se. Those interested in this area are referred to the writings of Hogan (1979) and Olson and Freeman (1979).

To provide an orientation on the environment that stimulated the current interest in occupational competency testing, this report begins with a review of competency-based vocational education and a discussion of evaluation concerns in vocational education. Next, some background information is presented along with definitions of a few key terms used in competency testing. After this, the current work of a number of organizations in the field of occupational competency measurement is reviewed. This is followed by a description of some recent developments in competency testing we feel merit special attention from test developers in vocational education. Finally, the technical and legal problems in setting standards of performance on competency tests are discussed.

Testing in general, and competency testing in particular, has been the subject of increased scrutiny in recent years. While considerable progress has been made, much remains to be done to overcome both the technical and social problems of such tests and to earn the confidence of the general public. As Klemp (1979, p. 52) noted recently:

> As we begin to develop tests that bear a more faithful resemblance to the competencies which they are designed to assess,

we may look forward to increasing use of these tests. More importantly, we may anticipate greater comparison and evaluation by an educated public of competency tests and their claimed predictive powers. This can only serve to make the testing movement more responsive to the needs of educators, employers, and society.

II. COMPETENCY-BASED VOCATIONAL EDUCATION

Whether one views competency-based vocational education (CBVE) as a great new movement, a new gimmick, or "just good educational practice" (Knaak, 1977, p. 40), there is little doubt that CBVE is rapidly spreading, and its effects on vocational education will be felt for many years to come.

There are many reasons why CBVE is attracting so much attention. According to Riesman (1979, p. 44), Americans are experiencing a renewed "fear of decay," in which they are worrying that the country "...has not only grown slack but is getting worse." Today, belief in one's own competence is no longer enough. Proof is demanded by employers, other gatekeepers such as graduate and professional schools, and consumers. Whether or not this pessimistic assessment is correct, the result is a new interest in approaches that give concrete evidence that individuals are able to do something well.

As evidence of this demand for competence, more than 70 percent of the states have passed legislation requiring some form of minimum competency testing of students (Knaak, 1980).

Definition

Performance-based, objectives-based, outcome-based, and competency-based are among the terms that are used interchangeably at times to label what this report terms competency-based education. The following definition seems to capture the main features of the concept:

> Competence-based education tends to be a form of education that derives a curriculum from an analysis of a prospective or actual role in modern society and that attempts to certify student progress on the basis of demonstrated performance in some or all aspects of that role. Theoretically, such demonstrations of competence are independent of time served in formal educational settings. (Grant, 1979, p. 6)

Hirst (1977, p. 32) ties the concept and definition to vocational education: "...competency-based vocational education is a systematic approach to instruction, aimed at accountability, based on job-derived standards, and supported by a feedback mechanism..."

Distinguishing Features of CBVE Programs

Beyond the label, what are the distinguishing features of CBVE? In general, CBVE stresses in-depth analyses and continuing adjustment to employment needs, coupled with the collection of student task-performance data as an aid in bringing student performance up to standard and for improving learning materials and instructor effectiveness. Hirst (1977, pp. 32-35) states that competency-based (or performance-based) vocational education can be broken down into the following basic components:

- Assessing available information to make realistic estimates of future employment opportunities
- Specifying the tasks that workers perform
- Conducting occupational surveys to determine task importance, task difficulty, and the experience level of workers performing the task
- Analyzing occupational survey data, documenting sub-tasks, and developing the performance objectives containing observable actions, situational conditions, and criteria for success
- Reviewing existing materials and media for applicability to the performance objectives
- Developing necessary new materials and media
- Preparing lesson plans specifying both teacher and student performance
- Testing the effectiveness of instructional materials, media, and lesson plans against student task performance
- Revising materials and media based on student performance data
- Reviewing and updating task analyses and instructional programs

These are the basic steps which tend to distinguish the developmental sequence of CBVE programs from other forms of vocational education. Once in operation, such programs continue to show major distinguishing features. Russell (1978, pp. 55-56) reported a survey of CBVE programs and noted that exemplary or model programs either encompass or are striving to encompass the following:

- Pre-testing students upon entry to determine the skills they already have as well as objectives that need to be achieved
- Allowing each student to proceed to subsequent instruction as soon as performance objectives are attained
- Providing an alternative method of instruction if a student does not achieve a learning task
- Recording student performance as each objective is achieved
- Placing greater emphasis on exit requirements (proficiency) than on entrance requirements
- Assessing students on the basis of competencies, i.e., criterion-referenced testing is used

Effects of Competency-Based Vocational Education

Most of the literature overwhelmingly supports the concept and practices of competency-based education. According to Grant (1979, p. 12), "...there is no question among leaders of the movement that competency-based education does lead to a net increase in societal competence, and this is one of the strongest dynamics driving competence reforms."

There appears to be general agreement that the use of a competency-based approach makes the job of evaluation much easier, particularly at the classroom instruction level. "Vocational teachers who conduct competency-based programs...are in a good position to appraise their instruction by focusing on its products..." (Erickson, 1979, p. 257).

Pratzner (1979) cites greater ease in program articulation as an additional positive effect of competency-based vocational education. As an individual moves through critical school-to-school, school-to-work, and work-to-school transition points, curricula based on well-ordered hierarchies of competencies ensure smoother movement at each point. Pratzner urges the careful use of job analysis as the technique for deriving the most appropriate curriculum. He outlines the steps in such analysis and recommends that all institutions offering occupational preparation programs in a district, state, or region cooperate in the design and conducting of these steps. Also recommended are student competency transcripts that report the objectives achieved and level of performance achieved.

In a joint report on ways of improving cooperation and articulation among vocational education delivery agencies, the American Association of Community and Junior Colleges and the American Vocational Association (1978, pp. 22-23) made the following recommendations:

> The U.S. Office of Education should, through the Bureau of Occupational and Adult Education, develop a data bank of competencies needed by individuals to enter or qualify for work in a broad range of occupations. Appropriate criteria for assessing whether or not the competencies are achieved should also be required.
>
> The U.S. Office of Education and the National Institute of Education should consider the development of guidelines for incorporating competency objectives into vocational education curricula, including the development of life skills. Such objectives should be geared to the various levels of education, e.g., secondary, postsecondary, and adult, as well as to the career goals of the individuals concerned.

Hirst (1977, p. 35) suggests some positive effects of CBVE provided the program is well-planned and implemented, including:

- A success-oriented atmosphere for learning, where success is measured by job-derived standards as opposed to competitive performance among students
- A new approach to vocational education where learning becomes the primary reason for instruction and time frames becomes less important
- A more professional approach to teaching, with positive feedback on the teacher's performance and materials used
- The development of successful performers who take on more responsibility for their own learning

Ingram (1980, p.47) cites several additional benefits:

- Students are more likely to master content.
- Students master prerequisite material before advancing to new material and can receive credit for competencies previously mastered.
- Students know exactly what to expect.
- Students have more personal contact with the instructor.

Not all views of CBVE are positive, however. Grant (1979) maintains that implementing a competency-based approach means more of an organization's human and financial resources will have to be spent on "middling" students, those who enter the institution less than adequately prepared for training. The more capable students will succeed on their own and need less contact with teachers. "...this heavy workload with middling students is one of the strongest sources of faculty resistance to competence programs" (Grant, 1979, p. 12).

Oen (1980) presents a number of potential pitfalls in the implementation of CBVE, e.g., the difficulty in finding sufficient time to spend with each student, the lack of appropriate resources, and the need for extensive staff time for developing materials and for testing individual students.

Some additional cautions are mentioned by Ingram (1980), including the difficulty of implementing CBVE on a large scale, the need for more work on the part of both the instructor and the student and the requirement for more and better instructional materials. According to Ingram too many students who need deadlines and a more instructor-oriented class fall victim to self-pacing.

Blank (1980) reviewed several studies of the use of CBVE approaches with students and notes that the findings of these studies are mixed. Some show the competency-based approach to be significantly better; others show no real difference between CBVE and the traditional approaches. Upon a closer look, however, the following findings seem to emerge:

1. Superficial modifications to programs (e.g., using objectives or taping lectures) under the guise of CBVE did not make much difference in student learning, while well-designed and well-planned approaches tended to enhance learning significantly.

2. Compared with traditional approaches, CBVE can produce benefits other than improved learning outcomes on the part of students e.g., shorter training time, greater opportunities for success, and more positive attitudes.

Blank (1980) also mentions several reasons why CBVE has not been implemented more widely than it has to date, including lack of substantial

data showing the superiority of CBVE; the existence of only a few successful, ongoing programs to observe; and the general lack of understanding of the concept on the part of key decision-makers.

The above summary of recent literature reflects the generally positive attitudes of vocational educators toward competency-based approaches, despite the absence of objective evaluation data. This does not mean, however, that vocational educators are necessarily in agreement on the definition of competency-based vocational education nor does it suggest that CBVE programs are by definition effective. Along this line, Grant (1979, p. 5) offers a most reasonable point of view:

> One cannot be 'for' or 'against' competence-based education any more than one can be 'for' or 'against' testing. Armchair dismissal or unqualified acceptance of competence-based education, as with testing, is likely to be wrong-headed. One has to ask: What kind of competence program? For what purposes? Under what conditions?

Selecting Competencies for CBVE

Wentling and Lawson (1975) discuss four previously developed taxonomies that are relevant to the classification of occupational competencies for measurement purposes. Four domains are covered: (1) cognitive, (2) affective, (3) psychomotor, and (4) perceptual.

Cognitive competencies and objectives include the intellectual outcomes of vocational education. They are divided into two major classifications: (1) knowledge and intellectual abilities and (2) skills. These in turn are subdivided into six levels of increasing complexity: recall of factual information, comprehension, application, analysis, synthesis, and evaluation.

Affective competencies include attitudinal and interest-based learning tasks. The hierarchy begins with receiving or attending and moves into the more complex areas of responding, valuing, organizing attitudes and feelings into an attitudinal structure, and finally characterizing or making the attitudinal structure part of consistent behavior patterns.

The psychomotor domain includes both simple and complex motor skills. At the lowest level is competence in perception followed by guided response, set, mechanism, and complex overt response.

Finally, according to Wentling and Lawson (1975), the perceptual competencies in CBVE include behaviors executed by the learner in the presence of various stimuli. The levels range from sensation to figure perception, symbol perception, perception of meaning, and perception of performance.

Block (1978) attempts to answer the question, "What is a competent school leaver?" In his model of competence, individuals make certain demands of their environment. Competence reflects their capacity to effect these demands. Competent individuals are able to manipulate their environment with their bodies and minds. When they are able to do this successfully, they also attain affective mastery of the environment as they gain feelings of self-confidence.

At the same time, the environment makes demands on individuals, requiring them to interact effectively in three kinds of settings, each of which is defined by particular adult roles. Some of these settings and roles are automatically assigned by society; others are optional and can be selected and aspired to by individuals. The final set are those that are invented or developed by the individuals themselves. According to Block (1978, p. 13), competent individuals are therefore those who "...possess particular motor, intellectual, and emotional competencies to handle the various intra-, inter- and/or extra-personal demands that each environment presents."

The literature on competency-based vocational education does not present detailed hierarchies of competencies, although some lists of potential competencies (or performance objectives) are available. For example, see The National Center for Research in Vocational Education Series, Performance Content for Job Training (Ammerman, 1977 a,b; Ammerman & Essex, 1977; Ammerman & Pratzner, 1977).

Two major interstate curriculum development consortia, the Mid-America Vocational Curriculum Consortium (MAVCC) (Benson, 1978) and the Vocational-Technical Education Consortium of States (V-TECS) have been actively involved in occupational task analysis and the development of performance objectives for use in CBVE. V-TECS has developed catalogs of performance objectives, which are available only to states joining the consortium (Kelly & Law, 1978).

A competency-based Model of Vocational Education Programs was developed by the Huntington Beach (California) Union High School District (1976). In addition to manipulative/technical skills required by the various occupations for which training is provided, the basic competency areas of communication, computation, comprehension, and coping are required. Competencies are ranked as necessary, desirable, and optional in terms of successful job performance.

The Canadian government, through Canada Employment and Immigration (1979), has developed "common" and "transferable" skills used by workers and supervisors in various occupations. One of the main purposes of the study, in addition to providing information for instructional design, was to assist students and employees in identifying as many occupations as possible for which they were qualified, the idea being that individuals could then move horizontally from one occupation to another as needed.

Based on data gathered from 1600 written questionnaires, the researchers identified 588 tool skills used in 131 occupations, which they clustered into 137 "skill classes." According to the authors, "The data from these studies clearly indicate that the skills used in the craft trades have more commonalities than differences" (p. 7). A series of cross-referenced charts is included for comparing trades and skills.

Most writers focus on the process used to identify the competencies in the first place. Block (1978) finds fault with prescribed lists of competencies, stating that many such lists of competencies are disseminated with little or no rationale or history as to how the competencies were chosen. Users are expected to "buy" the lists on the basis of faith, reputation of

the authors, or face validity alone. Block advocates using a broad base of "stakeholders" from the community to determine the competencies needed by the students of that community.

Minimums versus Maximums

The minimum competency tests required in many states can have the adverse effect of encouraging students to do only what they have to do to get by. Knaak (1980) warns that basing CBVE on bare minimums would mean the failure of CBVE altogether. On the other hand, CBVE, including competency testing, that is linked to "mastery learning" could be "...the foremost educational tool of the century" (p. 48).

Mastery learning, which is based on maximums, not minimums, was described by Bloom (1968) and drew heavily from the work of Carroll (1963) and others. Block (1971) later elaborated upon the concept and continued to collect evidence supporting its effectiveness. In essence, the mastery learning concept holds that most individuals can attain a high level of competence in most subjects (or skills) if they are given enough time and the right type of sequenced, individualized instruction.

An example of the mastery-based CBVE approach is the program at the Area Vocational Technical Institute in White Bear Lake, Minnesota (Knaak, 1977). Mastery levels are firm, and the school does not differentiate among "levels" of mastery. A student who does not complete requirements is credited with fewer masteries, not the same masteries at a lower level. Mastery is defined through the use of manufacturers' guides and rate books, program advisory councils, instructor experience, experience of students in intern programs, and follow-up studies.

Implementing a CBVE Program

Oen (1980) has outlined a series of practical planning and implementation steps for setting up a CBVE program; while a very detailed description of how to implement CBVE is contained in the five-volume series, Performance Content for Job Training, published by the National Center for Research in Vocational Education. (Ammerman, 1977 a,b; Ammerman & Essex, 1977; Ammerman & Pratzner, 1977). Finally, Walejko (1977) presents a clear description of

implementing CBVE within the classroom. Included are very practical suggestions for coping with the requirement for "being in thirty places at the same time" and ways of evaluating the approach's effectiveness.

Proponents as well as opponents of competency-based education offer a number of recommendations. For example, Grant (1979, p. 17) urges educators, institutions, and funding agencies to:

1. foster greater dissemination of what has been learned in the experiments undertaken thus far under the competence-based label

2. provide more opportunities for exchange of faculty and visits by faculty to successful programs

3. encourage the development of new modes of assessment and the further refinement of methods already developed...

4. sponsor more precise comparative studies of the long-range effects of competence-based programs as compared with more traditional forms...

In summary, competency-based educational programs in general, and vocational education programs particularly, are growing in popularity. The literature abounds with reports of positive effects of CBVE, and yet most authors cite a number of cautions that could develop into pitfalls unless the approach is well-planned and implemented. Instead of searching for the "ideal" taxonomy of vocational education competencies, the literature supports the idea of concentrating on careful development of processes for choosing appropriate competencies. Finally, the concept of "mastery" rather than "getting by on bare minimums" is urged as the criterion for CBVE programs, but implementers are urged to design programs that make it possible for all, or nearly all, students to succeed. To meet this goal without compromising program quality, vocational educators must be given the flexibility to adapt instructional strategies and to modify course length to meet individual student needs.

This movement toward competency-based education has been a major factor in the increased demand for objective measures of occupational competency. Another stimulus for better student competency measures has been the recognized need for improving the evaluation of vocational education programs, as discussed next.

III. EVALUATION CONCERNS IN VOCATIONAL EDUCATION

Since the passage of the Vocational Education Act of 1963 (P.L. 88-210), pressures have been mounting for evaluating the effects of vocational education. These pressures gained strength in the 1968 amendments to the Act (P.L. 90-576), which emphasized the requirement for evaluation at the state level and, in addition, directed the National Advisory Council on Vocational Education to conduct independent evaluations of programs carried out under the Act. Each state was also required to have a separate Advisory Council on Vocational Education, and these councils were encouraged to use funds for third-party evaluations of program effectiveness.

With the passage of the 1976 amendments to the Act (P.L. 94-482), evaluation of vocational education programs has assumed even greater prominence and visibility. Among the expanded evaluation requirements in these amendments are annual state accountability reports and follow-up studies. In addition, at least ten state programs are to be reviewed by the Bureau of Occupational and Adult Education, (now the Office of Vocational and Adult Education of the Department of Education), including an analysis of the programs' strengths and weaknesses and a fiscal audit of the programs. Of special importance is the provision that annual revisions of each state's five-year plan are to include information on how program evaluation data are being used by the state to improve its programs. The Commissioner is also required to make an annual evaluation report to Congress.

Continued Criticism of Vocational Education Evaluation

The quality of vocational education evaluations, especially those conducted during the 1960's and early 1970's, has been criticized repeatedly. Wentling and Lawson (1975) note that despite pressures for improvement, evaluation efforts seldom appeared to meet the intent of Congress. Abramson (1979) reports that most evaluation studies in the 1960's were narrow in focus and were generally limited to summative evaluation designs.

Stromsdorfer (1972) reviewed and synthesized a number of the cost-benefit analyses of vocational education done in the 1960's. He criticizes most of the findings on the grounds that they had poor designs and he urges

that issues such as impact of vocational education on values and preferences of students and the influence of unemployment rates on the determination of program costs and benefits be addressed in future cost analyses.

According to Abramson (1979), the National Research Council noted in 1976 that the literature describing the evaluation of vocational education programs was discouraging and yielded little useful information for vocational educators. The Council concluded that there were insufficient data to allow for a comprehensive evaluation of vocational education or its supporting research and development.

Over the years, many have speculated on the problems that vocational education has faced in meeting the expectations of Congress. The Fiscal Year 1979 Federal evaluation report on vocational education, conducted by the Department of Health, Education, and Welfare, frankly admits the difficulty. "Measurement problems and interpretation ambiguities make it difficult to characterize vocational education and its Federal support as either a success or failure" (U.S. Department of Health, Education, and Welfare, 1979, p. 496).

McKinney (1977) notes that although there is general agreement that evaluation of vocational education programs is necessary for better decision making, for various reasons a comprehensive systematic approach to program evaluation has been slow to develop.

Wentling and Lawson (1975) attribute part of the problem to the Federal government's failure to adequately define evaluation, and to the lack of Federal guidelines on how to conduct an evaluation. Lee (1977) is particularly critical of the quality of the basic vocational education statistics.

According to Datta (1979, p. 38), "the greatest problems for implementing vocational education evaluation requirements arise from limitations in the state-of-the-art." She takes issue with the continuing focus of vocational education evaluation on employment rates of high school graduates (which are influenced so heavily by economic conditions), rather than dealing with a wide range of occupationally related skills, or employability, over which educators have much more control.

Others have also taken issue with the narrow focus on employment of graduates as the main criterion of success. The Fiscal Year 1979 evaluation report of the U.S. Department of Health, Education, and Welfare states: "One problem (with using employment statistics) is that employment is not the primary objective of all vocational students. Another is that economic conditions probably much more powerfully influence employment among youth than curriculum choice" (U.S. Department of Health, Education, and Welfare, 1979, p. 496).

Morrell (1979, p. 242) agrees:

> Employment is not the only legitimate outcome variable for the evaluation of vocational programs. There is a psychosocial process to vocational training/rehabilitation, and evaluation information must be obtained on several relevant aspects of that process. Without such information one cannot know if a person's employment potential has been changed, or if the training has had significant secondary effects.

Venn (1979) argues that traditional evaluation criteria for vocational education are not relevant to the future, and that continued use of these criteria will decrease the effectiveness of vocational education. He suggests four criteria for consideration: (1) instructional and program quality; (2) program relevance to individual and societal needs in relation to work; (3) program impact on organization, policy, support, and use of vocational educat[illegible] and (4) individual transition to, and growth in, the work world.

Vocational Education Evaluation Begins to Mature

The 1970's saw the beginning of some improvements and promising trends in the field of vocational education evaluation. Denton (1973) outlines an evaluation model which includes needs assessment, development of philosophy, writing of objectives, statements of criterion questions, data collection, data analysis, formulation of recommendations, and decision making. Abramson (1979) comments that this model is praiseworthy as an important step in the synthesis of general evaluation literature and its application to vocational education.

Carbine (1974) goes beyond the usual variables included in cost-benefit analyses of vocational education programs as he urges evaluators to consider such issues as equity, socialization, and career decision making in cost-benefit analyses.

In 1978, Dunn developed an evaluation model for the directors of vocational education in the state of New York. His systematic model incorporates formative and summative evaluation procedures, including goal-free evaluations, to study vocational education processes used to produce end "products." Included in the model are three phases: planning, preparation, and operation, a systems approach similar to Denton's and other comprehensive evaluation systems used in both vocational and nonvocational settings.

Orlich, Anderson, Dodd, Baldwin, and Ohrt (1978) reviewed the literature and outlined procedures for (1) using planning as a method of evaluation, (2) evaluating short-term programs, and (3) using criterion-reference testing as an approach to evaluating student outcomes and instruction. This use of criterion (competency) based strategies has increased in importance.

By the late 1970's, evaluators were trying to incorporate more variables in their evaluation designs, as they tried to be more sophisticated in the measurement of the effectiveness of programs. For example, Darcy (1979) notes six components of a vocational education system that could be evaluated: institutional context, student characteristics, resources, program goals, educational processes, and learner outcomes. He presents a model and guidelines for evaluating each of these components.

Increased Emphasis on Competency-Based Learner Outcomes

Vocational education evaluation has always focused on learner outcomes, but most past efforts concentrated on the outcome of immediate employment as the real "bottom line" of vocational programs. Acquisition of employability competencies was considered a secondary benefit. As Darcy (1979) notes, the 1976 amendments were encouraging in that they provided another officially recognized outcome criterion in addition to that of employment. For the first time, states were to solicit employer assessment of how well

students were trained and prepared for employment, a measure of what students actually knew and could do.

A comprehensive review of studies related to vocational education outcomes was conducted by Taylor, Darcy, and Bolland (1979). In their annotated bibliography, 80 documents are summarized and cross-referenced. A review of the summaries indicates an increasing trend toward measuring student competencies (skills, knowledge, and attitudes) as part of a comprehensive evaluation effort.

The American Association of Community and Junior Colleges and the American Vocational Association (1978) support the trend toward competency-based learner outcomes. As noted earlier, their joint report includes recommendations that the U.S. Office of Education develop a data bank of competencies needed by individuals to enter or qualify for work. They also urge the U.S. Office of Education and the National Institute of Education to develop guidelines for incorporating competencies into vocational education curricula.

Datta (1979, p. 52) proposes a set of outcome criteria and potential measures "which seem consistent with the 1976 amendments." Competency-based measures form a major part of her recommendation.

In summary, evaluation of vocational education programs is becoming more complex, and the resulting data being generated are more plentiful than in the past. Trends that could mean an improvement in vocational education evaluation are:

- the new emphasis on comprehensive, systematic evaluation designs
- the addition of criteria other than immediate employment to measures of program success
- the growing interest in measurable learner competencies as part of an effective evaluation.

The following section takes a close look at the history of measuring occupational competencies and provides definitions of some basic terms in the field.

IV. BACKGROUND OF OCCUPATIONAL COMPETENCY TESTING

A Bit of History

The formal measurement of competency in task performance can trace its roots back to the very early history of the testing movement. The military, in both World Wars I and II, contributed significantly to the use of tests for personnel classification and performance evaluation. Chapman, in 1921, noted that one of the important outgrowths of World War I Army Personnel Research was the development of the trade test. This instrument was devised, in his words, "to make it possible for a trained examiner, unskilled in any particular trade, to measure in objective terms the trade standing of any recruit claiming skill in any of the several hundred trades necessary to the work of the Army" (p. v).

Chapman, a member of the Army Trade Test Division of the Committee on Classification of Personnel, defined "trade" very much the way "occupation" is defined today, and hence the term "trade test" was used synonymously with "occupational test" and "professional test." It encompassed such diverse occupations as those of surveyor, cook, statistician, and typist. Trade ability, according to Chapman (1921, p. 12) signified "what is commonly meant by a person's competency to follow a trade, occupation or profession." The concept of testing in Chapman's 1921 handbook was also much broader than has often been found in more recent test development guidelines and included the techniques and use of four test types: the oral trade test, the picture trade test, the performance trade test, and the written test.

World War II and the decade that followed gave another push to competency testing, with special emphasis being devoted to measuring the proficiency of aircraft pilots and equipment maintenance personnel (Flanagan, 1948). Glaser and Klaus (1962) provide an extensive review of this research and its applications, in their discussion of the problems of measuring proficiency of the human component in man-machine systems.

Today, the military still stands as one of the major developers and users of occupational competency measures. Their current activities will

be discussed later in this report, along with a number of other organizations that are active in the field today. While the number of these organizations is growing, much remains to be done before we can truly say that occupational competency testing has reached its full potential. In his sweeping criticism of the wholesale use of intelligence tests by schools, colleges, and employers, McClelland (1973) calls for competency testing as an alternative approach. As viewed by McClelland (p. 7), "the best testing is criterion sampling... There is ample evidence that tests which sample job skills will predict proficiency on the job." McClelland minces no words when he states: "Criterion sampling means that testers have got to get out of their offices where they play endless word and paper-and-pencil games and into the field where they actually analyze performance into its components."

In speaking about the trade testing movement in 1921, Chapman (p. vi) noted that the "movement is only in its infancy, but the methods that have been evolved will prove a firm foundation upon which an elaborate superstructure can safely be built." Nearly 60 years later, occupational competency measurement is still in its adolescence.

When one considers the time and the cost required to develop and validate occupational competency measures and the fact that performance measures typically require individualized testing and scoring, it is not surprising that the full utilization of competency tests has been slow. And yet, as Knaak (1977, p. 39) points out, the development and testing of criterion-referenced knowledge tests and performance checklists is a task of "monumental importance in the competency-based learning system." While Knaak stresses tests of technical knowledge and job performance together, such tests alone do not tap the full range of demands made upon graduates of vocational education programs.

Haller (1978) challenges the narrow assumption that students have to gain competency in meeting only the technical requirements of an occupation. He stresses the need to bring <u>people</u> back into competency-based curricula, rather than assuming that occupations are performed by robots. According to Haller (p. 35):

> It behooves us, then, to remember that accomplishment and success at work are not solely dependent upon technical skill. All work involves relations with others--peers, customers, clients and supervisors. It is the quality of these relationships that is crucial for success. It is others who can make our work sweet--or sour.

Considerable work has been under way recently in this area of affective work competencies or job survival skills (Beach, 1978; Kazanas, 1978; and Nelson, 1977).

Before discussing some of the organizations that are currently active in competency test development, a few definitions are in order.

Some Definitions

To set the stage for the remainder of the report, it would seem wise to review two terms being used widely and in various ways in referring to competency measures: criterion-referenced measurement and performance testing.

The term "criterion-referenced measurement" was introduced by Glaser and Klaus (1962) in their discussion of assessing human performance in man-machine systems. As defined by them, criterion-referenced testing depends on an absolute standard of quality and provides explicit information concerning what an individual can or cannot do, independent of the performance of others. In contrast, norm-referenced measurement (the well-known standardized achievement tests, for example) indicates the relative standing of individuals with respect to a given task. Glaser and Klaus (1962, pp. 421-422) explained further:

> Underlying the concept of proficiency measurement is a continuum of skill ranging from no proficiency at all to perfect performance. In these terms, an individual's efficiency at a given task falls at some point on the continuum, as measured by the behaviors he displays during testing. The degree to which his proficiency resembles desired performance at any specified level is assessed by criterion-referenced measures of proficiency. The standard against which an individual's performance is compared, when measured in this manner, is the behaviors which define each point along the individual skill continuum. When used in this way, the term 'criterion' does not necessarily refer to final on-the-job behavior. Criterion levels can be established at any point in training where it

is necessary to obtain information as to the adequacy of an individual's performance.

Shaycoft (1979, pp. 4-5), in her recent handbook, goes beyond this definition and divides criterion-referenced testing into two major categories: domain-referenced and objective-referenced tests. In domain referencing "the overall score has absolute meaning (criterion-referenced meaning) in the sense of indicating what proportion of some defined domain the examinee has mastered... The domain-referenced measurement is designed to yield a continuous score scale in which a maximum represents 100 percent mastery of any part of that domain." According to Shaycoft, an objective-referenced measure typically consists of a comparatively small number of items drawn from a larger set of possible items, deals with a specific objective, and "usually yields just a dichotomous score that indicates whether the examinee has reached the designated standard of performance corresponding to the specified objective."

On the other hand, Popham (1975) defined "criterion-referenced testing" in much the same way Shaycoft and many others have defined domain-referenced testing. Hambleton, Swaminathan, Algina, and Coulson (1978) agree with this definition; however, they note that if one accepts Popham's definition most of the tests labeled as "criterion-referenced" today should actually be called "objectives-referenced tests." Objectives-referenced tests, they note, consist of items matched to objectives but not considered a representative set of items from a clearly defined domain.

Suffice it is to say that enough confusion surrounds the definition of criterion-referenced testing, objective-referenced testing, and domain-referenced testing that one should make certain what definition an author or speaker is using. In this report, the more general definition of criterion-referenced testing, encompassing both domain-referenced and objective-referenced tests, will be used.

Some confusion also surrounds the use of the term "performance testing." Sanders and Sachse (1977) note that if either the stimulus (the test item or the problem presented) or the response has high authenticity

(fidelity or realism), the instrument can be labeled an "applied performance test." While Fitzpatrick and Morrison (1971) define a performance test as "one in which some criterion situation is simulated with more fidelity and comprehensiveness than in the usual paper-and-pencil test," it should be stressed that the types of tests (paper-and-pencil and performance) are not mutually exclusive. Boyd and Shimberg (1971), for example, describe a number of paper-and-pencil performance tests that can provide a direct measure of performance on certain tasks, such as developing site plans in the National Certifying Examination for Architects, the drawing of schematics for plumbing and electrical trades, and the reading of a vernier scale. While reading from a vernier taps only the ability to read the scale and does not reflect the full demands of the task of using a measuring instrument, Erickson (1976) cites the design of a computer program and the taking of a bookkeeping test as examples of paper-and-pencil tests that can serve as realistic performance tests for certain occupations.

As defined by the Clearinghouse for Applied Performance Testing (CAPT) (1980, p. 1), applied performance tests

> assess performance in real life, or simulated settings. Either the test stimulus, the desired response, or both, are designed to lend a higher degree of realism to the test situation than is the case with more traditional paper-and-pencil academic achievement tests. The identifying difference between applied performance tests and other academic achievement tests is the test's fidelity to the characteristics of a real-life task.

Two of the more common subdivisons of applied performance tests are work samples and simulations; however, the dividing line between these two categories may not always be sharp. For example, as noted by Fitzpatrick and Morrison (1971), any test may be viewed as a type of simulation. Nevertheless, work samples are generally defined as tests that employ an actual job situation, using the same tools and materials to perform some of the same tasks as that required on the job. Slater (1980) notes that "work samples have high fidelity to real-life tasks in the stimulus and response dimensions, but surrounding conditions tend to be somewhat artificial. Furthermore, even though the test stimulus mirrors that found in the actual workplace, it is in fact controlled and specified by the examiner, enhancing replicability of the task across examinees" (p. 8).

Simulations, on the other hand, require the examinees to pretend that they are engaged in some real task, the nature and context of which is described in some detail before the simulation begins. Simulation techniques can range from highly realistic tests, which tend to overlap work samples, to situations where considerable compromises are made in the stimulus and/or reponse dimensions in order to gain more control over the testing situation or to reduce costs. As viewed by Slater (1980, p. 10), simulation techniques (or situational tests, as they have also been called by Fitzpatrick and Morrison, 1971),

> cover the considerable middle-ground between objective paper-and-pencil examinations and work samples or direct assessment. Unlike the latter two types of performance testing approaches, which maintain high fidelity in the stimulus and response characteristics of tasks, performance tests labeled as simulations imitate, but do not duplicate, reality in these two dimensions. Of course, the conditions surrounding the simulated task are typically unlike those characteristics of real-life situations.

Included along the simulations are such varied techniques as role-playing, in-basket tests, management games, and leaderless group discussions, as well as paper-and-pencil problem-solving tests. Simulation techniques that merit particular attention are described in Chapter VI; but first a look at current activities in the area of occupational competency measurement.

V. CURRENT ACTIVITIES IN OCCUPATIONAL COMPETENCY MEASUREMENT

This chapter provides an overview of some of the more significant efforts currently underway, or recently concluded, in the area of occupational competency assessment. For purposes of discussion, these efforts are organized into three categories:

- Department of Education-Funded Organizations
- State Agencies and Consortia, and
- Job Performance Assessment in the Military.

Only major projects with broad implications for improving the methodology of competency assessment in vocational education are covered here. It should be noted that a great deal of additional work is underway today by commercial test publishers and other organizations engaged in developing proprietary tests or testing programs either under contract to business and professional associations or for direct sale to the public. These efforts are not described here.

Department of Education-Funded Organizations

Federal funds have supported directly several organizations concerned with vocational competency tests in recent years and, in addition, have contributed to a great many other test related projects through grant allocations to individual states. This section summarizes the current work of four Department of Education-funded organizations in the area of competency measurement: the Clearinghouse for Applied Performance Testing (CAPT), the National Center for Research in Vocational Education, McBer and Company, and the American Institutes for Research.

Clearinghouse for Applied Performance Testing (CAPT). Located at the Northwest Regional Educational Laboratory,[2] CAPT has been active in performance assessment since 1974 under the sponsorship of the National Institute of Education. One of CAPT's major activities has been the gathering of information on performance assessment and the preparation of bibliogra-

---

[2] 710 S.W. Second Avenue, Portland, Oregon 97204

phies, papers, and a periodic newsletter. Recently, CAPT published a series of 11 annotated bibliographies on the following specific topics in the area of applied performance testing: Performance Assessment Methodology; (Performance Assessment in) Reading, Writing, Speaking and Listening, General Problem Solving, Vocational Education, Experiential Learning, and Life Skills; and (Professional Competence Assessment in) Education, Managerial Skills, and Medical Fields.

CAPT also disseminates its information via training seminars and an annual conference. Beginning in 1980, CAPT has added a research function to its existing training and dissemination responsibilities. According to its May 1980 newsletter, CAPT activities will involve (1) synthesizing relevant research on given topics, (2) outlining research needs, (3) planning useful research strategies, and (4) conducting specific research projects.

National Center for Research in Vocational Education. During the past decade, what is now called the National Center for Research in Vocational Education has been active in vocational education research, development, program installation, and evaluation, with a large portion of its funds being provided by the U.S. Department of Education. Two of its recent efforts have been directed specifically toward occupational competency testing. In 1980, the National Center published a report entitled Performance Testing: Issues Facing Vocational Education, edited by Spirer. This report contains 16 papers encompassing four major issues for performance testing in vocational education: philosophical, technical, legal, and implementation. The range of disciplines represented by the authors and the breadth of their topics make this a stimulating book for those concerned with developing, implementing, or evaluating competency measures, especially those involving performance tests. Particularly noteworthy are the papers covering the legal issues, which discuss a number of lessons to be learned from the minimum competency testing movement.

A second recent effort of the National Center, which relates directly to occupational competency measures, is a series of six learning modules on instructional evaluation for assisting secondary and postsecondary vocational education instructors in designing and administering student evaluation measures in an occupational skills area. One module, for example, is

specifically devoted to skills assessment (1977). The six modules on instructional evaluation are part of a larger series of 100 performance-based teacher education (PBTE) learning packages for use in preservice or inservice training of teachers in all occupational areas.

McBer and Company. During 1980 McBer and Company completed an extensive review of current practices in assessing occupational competence, under the sponsorship of the National Institute of Education. The project encompassed a review of competence assessment practices in personnel selection in professional certification, and in higher education. Also included was a special study of assessment centers and their implications for education, as well as a review of how the courts have responded to assessment practices.[3]

American Institutes for Research. In October 1979, the American Institutes for Research was awarded a contract from what is now the Office of Vocational and Adult Education of the Department of Education to develop, field test, and disseminate comprehensive measures of competency in 14 selected occupational areas, and to design and help implement a program for continuing test development on a self-supporting basis. Tests are being developed for two occupations in each of seven curriculum areas:

- Agriculture
  - Agricultural Chemicals Applicator/Technician
  - Farm Equipment Mechanic
- Business and Office
  - Computer Operator
  - Word Processing Specialist
- Distributive Education
  - Food Marketing and Distribution
  - Hotel (Motel) Front Office

---

[3] The final report of the McBer study, edited by George O. Klemp, Jr. is available from the Educational Resources Information Center (ERIC). Because the report contains over a thousand pages, each section of the report is distributed separately by the ERIC Document Reproduction Service (P.O. Box 190, Arlington, Virgina 22210). The seven Clearinghouse numbers run from ED 192 164 through ED 192 170.

- Health
  - Dental Assistant
  - Physical Therapist Assistant
- Home Economics
  - Fashion/Fabric Sales and Sewing
  - Food Services (Front of the House)
- Technical
  - Electronics Technician
  - Water/Wastewater Technician
- Trade and Industry
  - Carpenter
  - Diesel Mechanic

The 14 competency measures, containing cognitive, performance, and affective components, are intended to serve two major purposes (1) to help teachers and administrators of secondary and postsecondary vocational education programs evaluate and improve specific areas of their vocational programs, and (2) to provide an objective basis for informing students, parents, and prospective employers about the progress made by students in acquiring specific, job-related competencies. A more detailed description of the project methods was presented recently by Chalupsky (1980).

State Agencies and Consortia

Described next are four occupational competency testing efforts with heavy involvement of state agencies at the present time. It should be stressed that the descriptions that follow are not meant to be exhaustive, nor should the reader assume that the organizations listed below are the only state agencies active in occupational competency testing today. Nevertheless, these four--the National Occupational Competency Testing Institute (NOCTI), the Ohio Department of Education, the Florida Department of Education, and the Council for the Advancement of Experiential Learning (CAEL)--do represent some of the more significant projects.

National Occupational Competency Testing Institute (NOCTI). For the past 15 years, NOCTI has been actively involved in the development and implementation of occupational competency measures. While the early start-up work was funded by a Federal grant (Olivo, Panitz, Schaefer, Nelson &

Barlow, 1973), what is now the NOCTI consortium is supported by its 46 member states and territories. Until very recently, NOCTI was concerned exclusively with teacher competency tests in selected trade and industrial occupations. The current series of NOCTI teacher tests consist of 47 examinations of which 26 were developed by NOCTI and 21, by the state of Pennsylvania. These tests can be used either as norm-referenced or criterion-referenced measures. Because in some instances there are two tests in the same occupational area, the 47 tests cover 38 different occupations. Table 1 presents the 38 occupations included in the NOCTI test series, as reported in the Fall 1979 edition of the NOCTI NEWS.

It should be noted that these tests are administered only by NOCTI area test centers and are not available for purchase or loan.

According to the 1979 NOCTI Technical Supplement, the examinations consisting of both written and performance tests have been standardized on populations of journeymen; national norms are updated regularly. The test purposes, as stated in this bulletin, are as follows:

(1) to measure the occupational competencies of skilled craftspersons who are interested in teaching in their occupational speciality,

(2) to verify occupational competencies as part of teacher certification requirements in certain states, and

(3) to promote a valid foundation for granting collegiate credit based on satisfactorily passing the written and performance tests for occupational competence.

As viewed by Olivo (1980), the results of the National Occupational Competency Testing Project for Teachers established the feasibility of forming a consortium of states for occupational competency. Nevertheless, Hoyman (1978), in discussing the problems of establishing a centralized program for administering occupational competency assessment, cited evidence indicating that the vast majority of occupational competency examinations adminstered by NOCTI in 1977 were conducted by only four of the states represented in the consortium.

TABLE 1

Occupations Covered by
NOCTI Teacher Occupational Competency Tests*

Air Conditioning and Refrigeration	Electronics Communications
Airframe and Power Plant	Industrial Electronics
Architectural Drafting	Machine Drafting
Audio Visual Communication	Machine Trades
Auto Body Repair	Major Appliance Repair
Auto Mechanic	Materials Handling
Brick Masonry	Mechanical Technology
Building Contruction Occupations	Painting and Decoration
Building Trades Maintenance	Plumbing
Cabinet Making and Millwork	Power Sewing
Carpentry	Printing
Civil Technology	Quantity Food Preparation
Commercial Art	Radio/TV Repair
Computer Technology	Refrigeration
Cosmetology	Sheet Metal
Diesel Engine Repair	Small Engine Repair
Drafting Occupations	Textile Production/Fabrication
Electrical Installation	Tool and Die Making
Electronics Technology	Welding

* Reprinted from Fall 1979 NOCTI NEWS

NOCTI has also become active recently in coordinating the development of performance measures for the Student Occupational Competency Achievement Testing (SOCAT) Program. During 1979, seven states (Alabama, Florida, Maryland, New Jersey, Ohio, Oklahoma, and New York) pooled their financial, material, and human resources for test development, as part of this consortium. A total of 20 tests are currently under development as shown in Table 2. These tests will be combined with paper-and-pencil measures developed earlier by the Instructional Materials Laboratory (IML) of the Ohio State University for the Ohio Department of Education's Vocational Achievement Test Program (1979). A third component, the California Short Form Test of Academic Aptitude, is also included in the overall testing package (Olivo, 1980).

TABLE 2

Student Occupational Competency Achievement Tests
Under Development by the NOCTI Consortium*

Accounting/Bookkeeping	Horticulture
Agricultural Mechanics	Industrial Electronics
Air Conditioning, Heating, and Refrigeration	Machine Trades
Auto Body Repair	Masonry
Auto Mechanics	Plumbing
Construction Electricity (House)	Practical Nursing Occupations
Drafting	Printing
Fashion Construction Services	Radio/Television
General Merchandising	Small Engine Repair
General Office	Welding

*Excerpted from NNCCVTE Newsletter, Fall 1980, VI, p.1.

According to Olivo (1980, p. 61),

> Currently, NOCTI plans to continue the developmental effort over two years. Each year, new tests in 10 occupation areas (cutting across several vocational fields) will be developed and field tested. At the same time, further analyses will be carried out for closer articulation of the written tests from the Ohio IML with the performance tests developed through the consortium.

Ohio Department of Education. The State of Ohio Vocational Education Achievement Tests are developed by the Instructional Materials Laboratory of the Ohio State University. The Ohio Vocational Education Achievement Test Program consists of specially designed, paper-and-pencil instruments to be used by secondary school teachers, supervisors, and administrators for the improvement of instruction. Since its beginning in 1958, the program has expanded beyond the trade and industrial education area and now also includes tests in agricultural, business and office, distributive, health, and home economics education. A total of 25 tests have been developed in this program, as shown in Table 3.

The Instructional Materials Laboratory of the Ohio State University has also been assigned the responsibility for publishing and distributing the tests and for the scoring, reporting, and analysis of test results. The tests are tightly controlled and are not available for purchase or for review. Rather, they are loaned out as part of a total package service, including test scoring and feedback of information. Percentile norms are provided for each occupation and grade.

In their early days, the tests were distributed only within the State of Ohio. Now, however, the service is made available to other states for a fee, charged on a per-student basis. All test administration is scheduled during the first three weeks of March. Total test administration time is approximately five to six hours, spread over three days. One hour is devoted to Level 5 of the California Test Bureau Short Form Test of Academic Aptitude (SFTAA), while approximately 2 to 2-1/2 hours on each of the two remaining days are spent on the occupational items. According to the Ohio test program report, "the academic aptitude test results will give the teacher an indication as to how the students are using their mental capacity in a particular vocational area" (Ohio Department of Education, 1979, p.2).

TABLE 3

The Ohio Vocational Achievement Tests*

AGRICULTURAL EDUCATION

- Agricultural Business (1980)
- Horticulture (1979)

BUSINESS AND OFFICE EDUCATION

- Accounting/Computing Clerk (1980)
- Clerk-Stenographer (1979)
- General Office Clerk (1980)

DISTRIBUTIVE EDUCATION

- Food Marketing (1979)
- General Merchandising (1980)

HEALTH OCCUPATIONS EDUCATION

- Dental Assisting (1971)
- Medical Assisting (1975)
- Diversified Health Occupation (1975)

HOME ECONOMICS EDUCATION

- Nursery School Teacher Aide (1979)

TRADE AND INDUSTRIAL EDUCATION (AUTOMOTIVE)

- Auto Body Mechanics (1979)
- Automotive Mechanics (1977)

CONSTRUCTION TRADES

- Carpentry (1980)
- Construction Electricity (1974)
- Heating, Air Conditioning and Refrigeration (1977)

ELECTRONICS

- Communication Products Electronics (1974)
- Industrial Electronics (1974)

GRAPHIC COMMUNICATIONS

- Commercial Art (1980)
- Drafting (1977)
- Lithographic Printing (1977)

METAL TRADES

- Machine Trades (1975)
- Sheet Metal (1964)
- Welding (1979)

PERSONAL SERVICES

- Cosmetology (1980)

* Excerpted from Ohio Department of Education (1979, pp. 3-6)

Florida Department of Education. In addition to its responsibility for developing the welding and the air conditioning, heating and refrigeration tests under the SOCAT consortium, Florida has under way its own program for eventually testing student occupational competencies statewide as a means for improving the State's programs of vocational education. The program entitled "Occupation Proficiency Performance Standards" is mandated by Florida statute for the purposes of educational accountability and "the identification of minimal competencies students must have in order to perform effectively in the occupation for which they are trained" (Agee, 1980, p. 62).

The Chief of the Florida Vocational Program and Staff Development Bureau, summarizes this program as follows:

> The goal is to establish statewide achievement measures, which will eventually result in state certification of competency achievement for the completers and/or leavers of Florida's vocational education programs. Test items and scoring, according to current plans, will be provided by the Vocational Division of the Department of Education. Test administration and materials will be provided by the local school systems, and monitoring will be done by local advisory committee members (Agee, 1980, p. 62).

At the present time, tests are being developed in 11 occupations as shown in Table 4.

Council for the Advancement of Experiential Learning (CAEL).[4] CAEL is an educational association of some 250 institutions of higher education and other educational organizations working toward fostering experiential learning and the valid and reliable assessment of its outcomes. Beginning in 1974 as a research and development project involving the Educational Testing Service (ETS) and a group of colleges and universities, CAEL received its initial funding from the Carnegie Corporation of New York and subsequent funding from the Ford Foundation, the Lilly Endowment, and the Fund for the Improvement of Postsecondary Education.

---

[4] Formerly called Cooperative Assessment of Experiential Learning, CAEL is located at American City Building, Suite 212, Columbia, Maryland 21044

TABLE 4

Florida

Occupational Achievement Tests

Under Development

AGRICULTURE	HOME ECONOMICS
Nursery Worker	Child Care Worker
Tractor Mechanic	Food Service Worker (Cook)
BUSINESS	HEALTH
Clerk Typist	Hospital Ward Clerk
Secretary	Nurse Aide
Stenographer	
DISTRIBUTIVE	INDUSTRIAL
Salesperson, Parts	Bricklayer (Construction)

One of CAEL's primary objectives is the development of alternative procedures for assessing prior competencies acquired in the work environment and helping colleges translate these competencies into college credit where appropriate. Among the 24 publications on experiential learning and the assessment of prior learning currently being distributed by CAEL are several monographs emphasizing the assessment of occupational competencies. For example, Knapp and Sharon (1975) discuss a wide range of assessment techniques; while Sharon's handbook (1977) presents a model for assessing specific work competencies, and includes prototype assessment instruments for three occupational areas.

CAEL's major focus has been on the assessment of two large classes of learning: that which is sponsored by an institution and typically is conducted off-campus, and that which is not sponsored by an institution and usually occurs prior to matriculation. Willingham (1977, p. vii) notes that "the intention of the CAEL Project was not to produce standardized assessment techniques or tests but to develop general guidelines and principles that could be adapted to local circumstances and to individual learning."

Experiential learning appears to be a major concern of both CAEL and CAPT. In a recent paper focusing on experience-based learning in public schools, Owens (1979) defined experiential learning as a process that includes both planned and unplanned experiences involving learners "in meaningful activities and relationships with others in the community"... "The learner is helped by another person to examine the meaning and implications of these experiences for his or her future growth" (p. 33).

CAEL's project director, Warren Willingham, while recognizing that CAEL's emphasis has frequently been on nonclassroom situations, emphasizes that all learning, in order to be most effective, should have both an experiential and a theoretical component.

> Experiential learning and its assessment often receive emphasis in traditional classrooms through special projects, research, laboratory exercises, and so on. Classroom learning tends, however, to place more emphasis on the theoretical component, partly from habit and partly because of the inherent limitations in the types of experience that can be usefully mediated in the school or college setting. (Willingham, 1977, p. 1)

Job Performance Assessment in the Military

During the past 15 years, the military services have undertaken major projects designed to improve the assessment of both individual and unit performance. Considerable modifications have taken place; in general, the shift has been toward a more performance-based system. While this shift has occurred in all three services, it is especially apparent in the Army. Accordingly, performance assessment in the Army will be discussed first.

Army. The Army's Skill Qualification Testing (SQT) system represents a major departure from its predecessor, the Military Occupational Specialty (MOS) testing program, and reflects the current Army emphasis on performance-based training and testing (Nieva, Myers, & Glickman, 1979).

The early work in establishing the concepts and techniques of performance-oriented training for the Army is discussed by Taylor and Staff (1975). This work affected training procedures used by the US Army and was part of the overall trend toward performance-based testing using the SQT system.

Maier and Hirshfeld (1978) present a detailed history of the Army's job testing program and trace the development of criterion-referenced job proficiency testing. In their discussion of the previously used Military Occupation Speciality (MOS) proficiency tests, which they term traditional achievement tests, they indicate that there was no "definitive logical correspondence between test items and specific job requirements" (p. 2). In the late 1960's and early 1970's the Army began to use performance-based training and testing in entry level courses. These courses were based on critical job tasks and criterion-referenced standards of performance. Because the Army had such success with this, a policy decision was made to change from the existing norm-referenced MOS tests to the criterion-referenced mode of proficiency testing called Skill Qualification Tests (SQT) based on the Soldier's Manual.[5]

The new SQT's are performance based and criterion referenced. The MOS tests previously used were norm based and were for the most part paper-and-pencil tests. The SQT's also differ from the previous tests in that the SQT's were derived from critical tasks which were spelled out in the Soldier's Manual for that MOS. They are criterion referenced with predetermined absolute standards which can be interpreted without respect to performance of others taking the test (Nieva et al., 1979).

The SQT's generally contain a "Hands-on component (HOC) and a Written component." The written component is in many respects similar to the older MOS tests. While the SQT was originally conceived as an effort to get away from paper/pencil tests, as of 1979, most tests are generally at least 90 percent in written form (Nieva et al., 1979). Time and cost considerations have been suggested as the major factors constraining the development of the HOC's. Some work discussed by Cockrell and Kristiansen (1978) pointed out the possible utility of less than full hands-on testing in which TV is used as the input for the performance testing.

One interesting aspect of the Skill Qualification Testing system is that it provides for specific examinee preparation for the test. Through a system of notices and practice exercises, those planning to take the test

---

[5] Presents the job duties expected of an incumbent in each job.

can prepare using what is called the "Notice." This document contains examples of the items, sections to be reviewed in the Soldier's Manual, and an announcement of the testing dates. Nieva et al., (1979) found that those who received formal "crash courses" in preparation for the test did better than those who did not receive the courses. They point out that, "if the SQT is intended to be a measure of individual abilities and general level of competence, the appropriateness of 'training for the test' becomes somewhat questionable" (p. 18). It would seem, however, that the more inclusive the SQT is with respect to the critical behaviors required on the job, the less likely this training for the test will result in higher scores than actually reflected in overall job performance. To the extent that all critical job behaviors are not tapped by the test, specific cramming for a subset of critical behaviors would be inappropriate, especially when the examinee knows which subset will be on the test.

With respect to test coverage and the use of the HOC, Nieva et al. (1979) suggest that for higher level skills, especially supervisory, the use of an organizational performance assessment center approach might be useful. Thus, the meaning of "hands-on" would change depending on the nature of the job duties. Certain jobs or job duties might require direct observation of interpersonal behavior as the "performance test."

A complete handbook for the development of Skill Qualification Tests has been prepared by Osborn, Campbell, Ford, Hirshfeld, and Maier (1977). This handbook discusses in detail the construction of the hands-on component and gives step-by-step procedures and checklists. It also covers the development of scoring procedures for both process and product outcomes and the factors to be considered in identifying job elements that cannot be scored from the product. The steps suggested for developing scoring procedures for "processes" are:

- specify performance measures
- break out elements into actions
- eliminate non-necessary actions
- define error tolerances-accuracy requirements

- specify safety considerations
- specify time limits
- specify sequence of actions

For developing scoring for "products," the following are suggested:

- define acceptable product
- define observable standard for each dimension of the product
- specify time limit
- specify tolerances for each standard
- prepare scoring aids if they are appropriate or can be used
- prepare procedures to ensure that product is preserved

Osborn et al. (1977) also discuss methods and procedures for the validation of hands-on tests. The procedures suggested include the assessment of the tests, selection of experts, testing the experts with the test, obtaining experts' judgments after they have taken the test, and checking on the scoring agreement of the scorers.

As part of the literature and documentation available about the SQT's, a manual has been prepared to provide guidelines for the administration of the SQT's. This manual (Ford, Campbell, & Harris, 1976), while it contains a great deal of information specific to the military, also provides guidelines appropriate to other settings as well. Advice is given in areas related to the development of scorer training and materials, selection of test locations, determining equipment requirements, time requirements, use of scorers, and administration of the hands-on tests and written tests. It also provides tips on scoring and processing of the completed scores.

Taylor and Vineberg (1975) critiqued Army performance tests in use in 1975 as part of their work in developing guidelines for performance test development for the then new SQT's. They reviewed and selected existing tests at the time (not the SQT's) to demonstrate problems encountered in test development and prepared a summary of the problems they found. The

summaries, intended to be used as aids by those involved in the preparation of the SQT's, grouped the problems into the following categories:

- instructions to test administrator
- instructions to the examinee
- task boundaries (incorrect task limits)
- cues (overcueing)
- verbal substitutions for performance
- lack of realism in alternative solution provided
- mismatch of test objective and test content
- standardization in administration
- scoring procedures (lack of details about scoreable elements)
- use of technical manuals (standardized use of job aids)
- adequacy of sample of performance

In a discussion about the difficulties of developing and administering large-scale hands-on tests, Maier and Hirshfeld (1978) point out that the Army had to abandon the idea of extensive, large-scale, hands-on testing and chose to have a hands-on component only as a subset of the SQT, rather than being the major portion. In addition, because of implementation problems, an alternative form of hands-on testing, performance certification, was decided upon. The performance certification covers tasks that are too long or complex to be in the hands-on test and which are not appropriate to test in the written mode of the SQT. It is completed by the supervisor in the normal job setting. While the performance certification was developed to solve problems with the large-scale hands-on testing, it too has problems which have resulted in its remaining only a small portion of the SQT. These problems were related to, (1) standardization of job testing conditions across individuals, and (2) standardization of scoring by supervisors.

The validation of both the written and the hands-on tests was discussed by Maier and Hirshfeld (1978). They view validity with respect to the SQT as focusing on consistency; that is,

- consistency between content of the tests and job tasks
- consistency among expert reviewers of the tests
- consistency of the tests in identifying competent job incumbents

Thus, they suggest that tests such as the SQT be subjected to both content and concurrent validity checks.

Air Force. The US Air Force has devoted much attention to the development of aircrew performance measurement systems. As discussed by Waag and Knoop (1977), however, as of 1977 R&D efforts within the Air Force focused primarily on the development of performance measurements for use in laboratory research programs. Such work involved the extensive use of simulators, with the Advanced Simulator for Pilot Training (ASPT) being one development.

Work described by Christal (1974) has been directed toward developing job inventories for a variety of airmen duty positions. The approach involved the preparation of task lists using supervisors and technical school instructors, and then having workers check the tasks that they performed in their jobs. In connection with this major effort, an elaborate and complex computer analysis system was developed (CODAP). This system is used by the Air Force and by other military services to analyze, organize, and report occupational information. Procedures were developed which permitted the assessment of task difficulty (defined in terms of relative time to learn the task). Task difficulty indexes would be useful, according to Christal in determining training requirements. As part of the training requirements work, he described plans for developing procedures for evaluating tasks on factors such as:

- consequences of inadequate performance
- probability that a task can be performed without specialized training
- probability that a task will need to be performed in an emergency
- cost of including the task in formal training vs. on-the-job training.

The Air Force has also done considerable work measuring the ability of maintenance personnel to perform the key tasks of their jobs with the objective not only to develop a model battery of performance tests on electronic maintenance tasks but also to develop a series of paper-and-pencil symbolic substitute tests (Foley, 1975). It was concluded that the symbolic tests, for the most part, showed sufficient promise to justify further consideration and refinement. However, it was pointed out that valid symbolic substitute tests cannot be developed for any job activity until good job-performance tests are available (Shriver & Foley, 1974).

Navy. The US Navy has for years funded developmental efforts designed to study electronic maintenance tasks and to develop various proficiency tests related to such tasks. Included in these efforts have been tests for preventive maintenance, corrective maintenance, trouble shooting, and test equipment testing (Anderson, Laabs, Pickering, & Winchell, 1977). Recent programs have included the Personnel Readiness Training Program (Anderson, et al., 1977). In this program, performance-oriented tests were used to diagnose deficiencies in job performance among Fleet personnel. Included in the program were also self-instructional materials which could be individually prescribed to correct the identified deficiencies. The testing was conducted at central locations using specially equipped vans rather than on-site observation and testing aboard ships.

In discussing the practical considerations related to widespread highly technical testing utilizing special equipment, Anderson and his colleagues (1977) suggest that central testing is better than decentralized for the kinds of tests used in their study. Some of the reasons put forth are unique to the Navy situation in effect at the time of their study; however, others such as giving each worker the opportunity to perform a standard task under standard conditions generally ruled out performance on a task (whether Navy job related or not) in a natural setting, unless the task under consideration is one performed on a very frequent basis. While it may not be necessary to perform the task in a central location, it is essential that the specifics of the task be so documented as to provide standard conditions in different sites.

In preparation for their work on the development of a system for obtaining and reporting Navy job-performance capability, Pickering and Anderson (1976) conducted a survey and review of performance measurement literature. Their review led them to suggest that the Navy adopt an assessment system based on the quality control approach used in industry. Also proposed was a series of studies to provide additional information. In a later document, some results are reported by Bell and Pickering (1979). They note that, "in order to ensure Fleet readiness, the US Navy is continually seeking new approaches for assessing the job performance of its personnel" (p. vii). While many partial performance measurement systems exist, they report that the Navy does not have a comprehensive system for the measurement of job performance proficiency. In response to this need, a Performance Proficiency Assessment System (PPAS) was proposed. The approach used in the effort reported on by Bell and Pickering made a clear distinction between the concept of a PPAS and other performance measurement processes. The system they were developing would not be concerned with evaluating individuals or Navy units, but rather would be more like an industrial quality control method or procedures. In this approach, "relatively small samples of a product are tested periodically; and when deficiencies are found, appropriate corrective actions are taken" (p. 1). As a result of their study using advanced ASW team training exercises, they propose general requirements for a PPAS Data Collection System.

In their work on the Navy's Performance Proficiency Assessment System, Laabs and Kissler (1978) evaluated the reliability and validity of expert judgments of relative task criticalness that were obtained using Q-sort methodology. Validity of the judgments was checked by (1) intercorrelating the mean rank order of the performance domains yielded by this sorting process and the mean rating of consequences of improper and delayed performance, and by (2) having judges form an impression of each of several pairs of hypothetical sonarmen described in terms of their capabilities and then selecting one from each pair who they felt would contribute most to the sonar gang. Laabs and Kissler concluded that

> This application of the Q-sort methodology yielded reliable and valid expert judgments, and this methodology should be considered as an alternative to the more traditional rating

> scales. The card-sort technique, coupled with the appropriate clustering of job tasks, provides a method for the accurate identification of critical tasks. (1978, p. vi)

McCormick and his associates have done considerable work for the Navy on job characteristics and job dimensions. (See McCormick, Jeanneret, & Mechan, 1972.) A key element of this work was the developing of the Position Analysis Questionnaire (PAQ), which is a structured job-analysis instrument. According to McCormick, et al. (1972), the primary frame of reference underlying the PAQ followed, "the information-input, mediation, and work-output model...there being individual job elements relating to each of these." There were, in addition, job elements relating to interpersonal activities, nature of work situation and other miscellaneous aspects of work. In general, worker-oriented elements (used on the PAQ) are those "that tend more to characterize the generalized human behaviors involved" (p. 348). On the other hand, "job-oriented elements are descriptions of job content that have a dominant association with and typically characterize the technological aspects of jobs and commonly reflect what is achieved by the worker" (p. 348). Results indicate that "aptitude requirements and rates of pay can be predicted reasonably well from such quantified job-analysis data, thus suggesting that conventional test-validation and job-evaluation procedures might sometimes be eliminated" (p. 347).

VI. SOME RECENT DEVELOPMENTS IN OCCUPATIONAL COMPETENCY TESTING

The field of occupational competency measurement, while possessing a long history as mentioned earlier, cannot lay claim to a very long list of achievements. Nevertheless, there are several areas that merit some discussion here, especially in terms of their potential contribution to future competency test development efforts. These are: simulations, adaptive paper-and-pencil tests, confidence testing, and Rasch modeling.

Simulations

As discussed previously, the use of simulation techniques in competency measurement falls somewhere between objective paper-and-pencil tests and work samples or on-the-job checklists (Fitzpatrick & Morrison, 1971). Indeed, the greater the simulation fidelity or realism of the test, the more it tends to overlap work-sample techniques. When the term "simulation" or "simulator" is mentioned, the general tendency is to think of elaborate military training devices (such as the Link Trainer) or possibly the stress interviews that were carried out by the OSS to select agents during World War II (Office of Strategic Services, 1948). However, there are a variety of other simulation techniques that, while perhaps not quite as dramatic, are more likely to be included in occupational competency assessment batteries. Three of the more popular simulation strategies today are in-basket tests, management or business games, and leaderless group discussions. Even these techniques are used much more with administrative or management levels rather than with entry-level occupations, which are the prime focus of this project. Consequently, only brief summaries of these techniques will be included here.

In-basket tests. As the name implies, the examinee in this test is confronted with an assortment of reports, memos, incoming letters, telephone messages, and other items that have supposedly collected in the in-basket of the job she or he is applying for. The candidate is instructed to handle all of these documents (each of which can be considered a test item) in the most appropriate manner. While typically each action produces a written product for later scoring, the completion of the in-basket action is often followed by an interview or questionnaire wherein the examinee justifies

the actions taken. While this testing approach was originally designed for selecting managers and is considered one of the most popular situational exercises in managerial assessment centers (Finkle, 1976), it has also been used to assess the capabilities of school principals, police officers, and military personnel (Knapp & Sharon, 1975).

Management games. Business or management games can trace their origin to the war games used for training in the German army in the early 19th Century (Fitzpatrick & Morrison, 1971). Participants in a business game are typically instructed to operate a business, either cooperatively or competitively, and are faced with making a variety of decisions as problems arise. The consequences of their actions, in relation to the moves and countermoves of their competitors, are fed back to the players, especially in those games that are computer-based. Knapp and Sharon (1975, p. 16) categorize games as:

1) media-ascendent simulation games mediated by machines;

2) interpersonal-ascendent simulation games in which decision making, role playing, and player interaction are emphasized; and

3) non-simulation games which merely provide a competitive context for learning concepts and principles.

The game environment can serve as a vehicle for evaluating not only the technical aspects of managerial performance but such factors as interpersonal skills, leadership skills, and organizational ability. Fink, Wagner, Behringer, and Hibbits (1974) note that games could be useful devices for assessing the skills of individuals during their professional training. However, in their state-of-the-art review of techniques for measuring complex behaviors, they found a reluctance to use games as assessment devices, possibly because of the high cost in both design and operation. In contrast, Knapp and Sharon (1974 p. 17) state that

> academic or management games could be one of the most economical assessment techniques, in that virtually hundreds of them are available through companies that specialize in manufacturing educational products. There is no evidence that 'canned' games make a difference in eliciting the desired behaviors; therefore, it is probably not necessary to develop tailor-made games for each assessment situation.

Leaderless group discussion. Another technique that has been found useful in evaluating interpersonal skills and leadership qualities is the leaderless group discussion. Here the participants are typically given a discussion question or problem dealing with supervision or management and told to arrive at a group decision. In one type of leaderless group discussion, roles are not assigned and the role that a person assumes is presumably the role that would be adopted in a natural group. In the case where roles are assigned, each participant is given a point of view that must be sold to the rest of the group. During the course of the discussion, evaluators observe and rate each group member on whatever criteria have been established beforehand. In their appraisal of this strategy, Knapp and Sharon (1974) note that the ability to work in a group situation is difficult to judge validly by such techniques as questionnaires and paper-and-pencil tests. Leaderless group discussions, they feel, might fill this assessment gap as "the most economical work sample of group behavior" (p. 15).

Adaptive Paper-and-Pencil Tests

One type of simulation that deserves to be discussed as a separate topic is known variously as diagnostic problem-solving tests (Fitzpatrick & Morrison, 1971), adaptive paper-and-pencil tests (Fink et al., 1974), and written simulations (McGuire, Solomon, & Bashook, 1976). One of the earliest reported paper-and-pencil approaches to testing problem-solving performance was the "tab" test developed by Glaser, Damrin, and Gardner, in the early 1950's (Fitzpatrick & Morrison, 1971), as a means for evaluating the efficiency with which a maintenance technician used information in diagnosing an equipment malfunction. The name, tab item or tab test, came from the fact that the examinee tore off a perforated tab each time he chose to carry out a specific action. Behind the tab were presented the results of each action. Based on the assumption that troubleshooting is most effective when the amount of wasted motion and of trial and error are kept to a minimum, scoring of the tab test was based on the number (and sometimes also the sequence) of the tabs that were pulled. Boyd and Shimberg (1971) describe an economical alternative to the tab test, which consists of writing the diagnostic action on the outside of a sealed envelope with the consequences on a slip of paper inside the envelope.

They point out that tearing open the envelope (in effect) gives the same irreversible evidence as does pulling a tab.

A later refinement of the tab test used an opaque, erasable ink printed over the answers previously hidden by the tab. The examinee removes the overlay by rubbing it with a pencil eraser. Another variation uses a special invisible ink for the outcome of each diagnostic alternative. The examinee brushes the area next to the option chosen and immediately learns what would happen as a result of that action. The National Board of Medical Examiners has gained over 20 years of experience with the use of patient-management problems (PMPs) involving written simulations for evaluating the clinical competence of interns in the National Board Examination. This test, which leads to certification for licensure, was the direct result of research on the dimensions of clinical competence conducted in 1959-60 by the American Institutes for Research (1976) applying the Flanagan critical incident technique (Flanagan, 1954). The PMP test format has also been adopted by a number of specialty boards for use in their certifying examinations. In a recent paper covering a variety of different simulation strategies, McGuire (1979) stated that patient-management problems in a paper-and-pencil format (using latent image or opaque overlay techniques for feedback) are now available for either individual or group administration and are amenable to computer scoring analysis.

One of the questions raised concerning patient-management problems is just what is it that they measure. Gampert (1980) conducted a study of patient-management problems administered to candidates as part of the licensing process for veterinarians, in an attempt to determine whether written simulations were measuring something other than factual knowledge. He correlated the scores of the National Board Examination for Veterinary Medical Licensing (NBE), a 435-item, multiple-choice test with a series of linear clinical simulation problems called the Clinical Competency Test (CCT). The CCT included 13 linear patient-management problems in three major content areas: small animal, food animal, and equine practice. He found that the product-moment correlations indicated the common variance between NBE and CCT was generally less than 10 percent. In conclusion, Gampert (1980, p. 10) notes

> the results of this study indicate that the CCT and NBE are distinct examinations assessing different aspects of veterinary medicine. Further, the results support the notion that the CCT is assessing case-management skills beyond those relating solely to problem content (e.g., small animal). The tentative conclusions of this study are that cognitive processes relating to different case-management skills are what the CCT is assessing. However, what these specific cognitive processes are (e.g., information processing) could not be determined from the data obtained.

McGuire (1979, p. 22), in summing up one of the areas of benefits in the use of simulation, stated that "all candidates can be allowed full responsibility to careen down their merry way to disaster without any risk whatsoever to anyone or anything other than a piece of paper, a computer, or the psyche of an examinee." As vocational education graduates become involved with progressively more complicated and expensive equipment, it is quite likely that simulations, and particularly written simulations, will achieve much greater prominence.

Confidence Testing

One "old" development that has been around for many years, but still relatively limited in its usage, is a strategy for weighting item responses to reflect the confidence that the examinee has in the correctness of each item response. The procedure, referred to as confidence testing, is designed to provide a maximum amount of information from a given set of items. Probably the simplest and most efficient scoring system in confidence testing is to ask the examinee to select the correct answer and then indicate how confident he or she is in the correctness of that response (e.g., "very sure," "pretty sure," "not sure," etc.). A more complicated scoring technique is to ask the examinee to distribute a number of points reflecting the confidence in each alternative that is presented in the test item.

While this technique has not been used to any great extent, except in research conducted on Air Force technical training, Fink, Wagner, Behringer, and Hibbits (1974, p. 25) cite the following advantages that have been claimed by proponents of the method:

1. Confidence testing may provide more information about an examinee's state of knowledge than does standard multiple-choice or other paper and pencil tests.

2. The instructor can use the information obtained from confidence testing to prescribe instruction tailored to the needs of each individual student.

3. The use of confidence testing requires each examinee to be more careful before responding to each question. This supposedly increases their sensitivity to the content of the questions.

The Rasch Model and Latent Trait Theory

One of the more serious problems facing those concerned with administering occupational competency measures is finding sufficient time to do a reasonably competent job of testing. As noted by Oliver (1978), "in most cases time will not permit the teaching and measuring of all the tasks that make up a job" (p. 48). Similarly, Abramson and Vogrin (1979, p. 374), in discussing the development of ISSOE (Instructional Support System for Occupational Education) in the State of New York, stress the importance of generating items for a reduced set of objectives and point out that otherwise,

> the need for testing each student as each major objective is completed will in fact make this system unmanageable. For example, if during the course of a school year, a teacher were to work with forty students and fifty major objectives, and if the students were to complete the fifty objectives, the teacher would be required to administer 2000 assessments--a situation which would require more than ten assessments per day for each school day.

A possible solution to the testing time problem may be found in the application of one or more of the models that have grown out of latent test theory (or item characteristic curve theory) developed nearly 30 years ago (Lord, 1952). The role of latent trait models in reducing the number of items that must be administered to a single individual (along with the crucial assumption of test homogeneity) is summed up by Hambleton and Cook (1977, p. 90):

> Given a set of test items that have been fitted to a latent trait model (that is, item parameters are known), it is possible to estimate an examinee's ability on the same ability scale from any subset of items in the domain of items that have been fitted to the model. (Of course, the domain of

items needs to be homogeneous in the sense of measuring a single ability. If the domain of items is too heterogeneous, the ability estimates will have little meaning.) In fact, regardless of the number of items administered (as long as the number is not too small) or the statistical characteristics of the items, the ability estimate for each examinee will be an asymptotically unbiased estimate of true ability, provided the latent trait model holds. Ability estimation independent of the particular choice (and number) of items represents one of the major advantages of latent trait models.

These models, representing rather complex test theory and advanced mathematics, have only recently begun to gain attention as practical tools. However, Hambleton and Cook (1977) note that a number of psychometricians feel test practitioners must become aware of the importance of these models. Their potential contributions, in addition to the saving of considerable testing time, have obvious application for maintaining test security inasmuch as different sets of items can be used for different examinees. Other benefits attributed to latent trait models include the development of tailored tests, where examinees take only those items that are matched tc their ability level, developing parallel forms of a test, and equating scores on two or more tests that measure the same ability.

According to Lord (1977), mastery testing also presents a particularly appropriate application of what he now refers to as "item response theory," since such tests come close to being unidimensional. He also notes that conventional testing with heterogeneous groups is not able to measure accurately both high and low ability levels at the same time. According to Lord, if we wish to measure accurately across a very wide range of ability, "we need to match difficulty level of the items administered to the ability level of the examinee tested. This is individualized or tailored testing" (1977, p. 125). While each examinee usually takes a different set of items, latent trait (or item response) theory presumes that the examinees can all be placed on the same scale.

Urry (1977), recognizing the successful application of latent test theory at the Civil Service Commission (now the Office of Personnel Management), states that widespread application of computer-assisted or tailored testing in the future was inevitable. In noting the dramatic reduction in

the number of items required for valid measurement, he reports that a recent analysis placed the cost of tailored testing at less than that of conventional paper-and-pencil testing, with a potential utility (dollar benefits adjusted for cost) far exceeding the conventional test battery. Other benefits cited by Urry include shortened testing time because fewer items will be necessary; immediate test-score reports, weighted for relevance to various occupations; reduction of possible test bias due to standardized administration; a lower risk of compromise because the test items will be displayed on computer terminals rather than being printed; and improved scheduling of examinations where it is necessary to test large numbers of candidates.

According to Wright (1977) use of the Rasch model simplifies the implementation of tailored testing. This model is the simplest of all the latent trait models, requiring only one item parameter, difficulty, to be estimated. All items are assumed to have equal discrimination power. In appraising the effectiveness of the Rasch model, Hambleton and Cook (1977) indicate that the assumption of all item discrimination parameters being equal is restrictive and that evidence is available suggesting that unless test items are specifically chosen to have this characteristic, the assumption will be violated. The model has also been challenged in its capability for equating tests vertically. Lloyd and Hoover (1980, p. 192) summarized their study as follows: "While latent trait methods show a great deal of promise for improving the horizontal equating of tests, the results from the present study, and others, indicate that the use of the Rasch model in vertical equating should be approached with extreme caution." On the other hand, Slinde and Linn (1979) indicate that while the model may not be adequate for extreme comparisons, it may still be useful where differences are less extreme and guessing is minimized.

While it appears that the Rasch model is closer to practical application than the other latent trait models, there is need for considerable research and experimental trial of all models before widescale application can be expected. Nevertheless, the possible payoff of these experiments for improving the efficiency of occupational competency testing would seem to justify such efforts.

VII. SETTING PERFORMANCE STANDARDS ON COMPETENCY TESTS

One of the areas in vocational competency testing that has not received sufficient attention is the setting of proficiency standards and the translation of these standards into specific test cut-off scores. This lack is especially noticeable in the case of criterion-referenced tests, where standards are assumed to reflect absolute skill requirements independent of the performance of other individuals. One can speak rather glibly of mastery learning in competency-based vocational education and yet not come to grips with the problem of specifically how to decide when vocational education students are competent to perform adequately on a job, based on their performance on a set of test items. While the problem is more noticeable in the case of paper-and-pencil items, it nevertheless exists for performance test items as well. In fact, in the case of performance tests, setting scoring standards may well be viewed as deceptively simple. Erickson and Wentling, in their handbook for teachers (1976, p. 349), tend to reflect the attitude that setting standards is a rather simple process. They cite a variety of situations, ranging from barbering to airline pilot programs, where the

> level of mastery is arbitrarily set at some level beyond a recognized level of minimum usefulness, but not at a level that is beyond practical necessity. That is, most office education teachers recognize that being able to type only 30 to 40 words per minute is generally not fast enough in the business world. However, to insist that each of their students be able to type 100 words per minute is beyond practical expectation in all but very specialized assignments. Therefore, 65 words per minute has been somewhat arbitrarily established as an acceptable level of mastery for first-year typing students. Similar thought processes are used to establish levels of mastery for students enrolled in most all occupational programs.

They note further (p. 400) that "the most recognized reason for determining levels of mastery and identifying deficiencies in students' abilities to perform at or above these levels is to serve as a basis for assigning student grades."

The view that choosing cut-off scores for performance tests is a relatively straightforward and simple process is also advanced by Livingston in

a report entitled Performance Testing: Issues Facing Vocational Education, which was published by the National Center for Research in Vocational Education in 1980. He states,

> Probably the most meaningful type of judgment is the direct judgment of examples of performance as acceptable or unacceptable. In most other kinds of testing, it is difficult to get meaningful overall judgments of students' performance; in performance testing it is easy. Judges' standards will vary, but these differences will tend to 'average out' if several different judges participate in the process. By analyzing the students' test scores together with the judgment of their performance, we can estimate the probability that a student with a given test score would be judged (by a randomly selected judge) to have performed acceptably. (p. 87)

Klein, in this same report, takes a more realistic view of the problem on setting cut-off scores and recognizes that such standards may need to be defended. He states, "Once such a cut-off point has been established, the results of the examination should be monitored. This will ascertain whether or not the measures are providing weights for meaningful decision-making. It is only through a constant reappraisal that appropriate cut-off scores can be maintained" (p. 79).

Klein also makes the case that test scores must be fair to both the examinees and to others who use the test results, and recognizes the limitations of test scores. He cautions, "The measures obtained from a performance test represent an estimate of an individual's performance under a given set of conditions. They cannot represent all of the characteristics required to perform a given task adequately" (p. 79).

Despite the rather casual attention devoted to cut-off scores in vocational education, they can present a variety of problems for educational institutions. These problems have been highlighted particularly in the recent experiences of minimum competency testing. Competency testing in the vocational area presents a number of parallels with the minimum competency movement. This was recognized by Bunda and Sanders (1979, p. 14) who, in editing a volume on competency-based measurement for the National Council on Measurement in Education, made the following introductory commentary:

> Although there is little to look at in the area of elementary and secondary education, there has been a relatively great amount of attention given to the measurement of competencies in professions such as nursing, law and medicine, and in many vocational areas. One reason for the great amount of activity in these areas has been the movement toward licensing and certification of members of the professions and trades. Those working in elementary and secondary education might profit from looking at their collective experience and by asking themselves, 'What can be learned from a review of areas with a relatively long history of research and development in competency measurement, and what does this history say as we consider adopting competency-based programs?'

On the other hand, Tractenberg (no date), in his discussion of the legal implications of performance testing in vocational education, recognizes that vocational educators can learn from the minimum competency movement. He notes, "Performance testing in vocational education has significant parallels to minimum competency testing in general education. Indeed, the momentum generated by the latter undoubtedly has contributed to the increased interest in the former; peaking of the minimum competency movement, or adverse court decisions, therefore, would have implications for performance testing" (p. 91).

The relationship between the minimum competency movement in general education and the measurement of skills of vocational education students is also reflected in the concern that the establishment of minimum standards may well become maximums or ultimate goals and limit the level of educational attainment. Shepard (1979) and Conaway (1979) discuss these concerns relative to general education; while Wentling (1980) raises this same issue in respect to the establishment of national or state standards for vocational competencies.

The recent attention that has been devoted to setting test performance standards or criterion levels relative to minimum competency testing suggests two categories of problems appropriate for consideration by vocational educators: technical problems in standards setting and legal problems.

Technical Problems in Setting Performance Standards

Occupying a notable part of the educational literature recently is a very basic question whether or not minimum performance standards can legitimately be established. The most vigorous condemnation of standards setting has come from Glass (1978). He views the attempts to establish all such standards, whether "mastery," "competency," or "proficiency," as "pseudo-quantification, a meaningless application of numbers to a question not prepared for quantitative analysis." For Glass, a tentative notion of "mastery...has been translated prematurely into the idea of cut-off scores and mastery levels. If ever there was a psychological-educational concept ill-prepared for mathematical treatment, it is the idea of criterion-referencing" (p. 242).

In his criticism of six classes of methods for setting criterion scores on criterion-referenced tests, Glass (1978, p. 246) goes well beyond the classroom usage of basic skills tests. He refers to the cutting scores on various licensing exams and notes that they have little to do with psychology and behavioral analysis.

> Written examinations for licensing automobile drivers have passing scores, usually at around 90% of the questions. Whether the number of errors permitted is 2 or 5 or 10 is arbitrary, and there is scant reason to believe that highways would be less safe if the permissible error rate on the test were doubled or tripled. ...These cut-off points have virtually nothing to do with defensible judgements of competent vs. incompetent.

While many have challenged the extreme tone of his criticism, Glass does indeed present a resounding alarm to awaken vocational educators as well as others who may have been lulled into lethargy by those who pass off the process of standards setting on competency tests as a trivial matter. While he supports the general idea of a continuum stretching from absence of a skill to "conspicuous excellence," Glass (1978, p. 251) disputes the notion that one can recognize the highest level of skill below which a person would not be able to succeed, whether in a trade, the next level of schooling, or general living.

> What is the minimum level of skill required in this society to be a citizen, parent, carpenter, college professor, key-punch operator? Imagine that someone would dare to specify

> the highest level of reading performance below which no person could succeed in life as a parent. And the situation is no different with a secretary or electrician--in case one wished to argue that minimal competence levels are possible for 'training,' if not for 'education.' What is the lowest level of proficiency at which a person can type and still be employed as a secretary? Nearly any typing rate above the trivial zero-point will admit exceptions; and if one were forced nonetheless to specify a minimal level, the rate of exceptions that was tolerable would be an arbitrary judgement.

Glass does not stand alone in his argument against criterion-referenced test standards. Burton (1978), for example, indicates that criterion-referenced testing became so important because its proponents felt that by directly "referencing" test performance to skills judged to be necessary they could avoid the difficult, expensive, and time-consuming process of correlating test performance with job performance and setting cut-off scores based on the probability of job success. In her view, "the original hope that criterion-referenced tests (with performance standards) would provide a simple or inexpensive tool for making decisions beyond the classroom level has not been realized" (Burton, 1978, p. 269). According to Burton, "there is something fundamentally wrong with the idea of using specific achievement measures when thinking about success or failure in any real venture" (p. 269). She argues that success can come from anywhere in the skill repertoire and that "performance standards are not sensible for any problem that has any more than a small, definable set of possible solutions" (p. 270).

A more tolerant, albeit critical, view of minimum test standards is voiced by Shepard (1979). After noting that standard-setting should be avoided whenever possible, and that pupils should rather be monitored along a performance continuum, Shepard goes on to suggest procedures for making standard-setting as "thoughtful and well-informed as possible" (p. 67). Several of her suggestions are particularly relevant for vocational education and, indeed, support some of the recommendations made by more enthusiastic supporters of test cut-off scores. For example, her recommendations that (a) setting standards be an iterative process and that (b) normative data on past test performance be considered in setting such standards support those of Hambleton (1978), Linn (1978), and Popham (1978).

Shaycoft (1979, p. 68) points out that norms inevitably enter the picture, consciously or unconsciously, whenever standards of competence are set. She cites the licensing of physicians as a clear indication of the impact of normative data.

> Physicians are certainly expected to know how to diagnose acute appendicitis, and to know the proper way of handling it. But they are not expected to know how to treat a color-blind person to give him normal color vision, or how to cure any disease for which at present there is no known cure. Not requiring candidates for licensure as physicians to know how to achieve these cures does not mean it would not be desirable for them to have these abilities; it merely means that they do not, and that this fact is recognized.

Shepard (1979) also suggests that all relevant audiences be involved in standard setting. This reflects Jaeger's concern (1979) that different samples of judges would set different standards, and goes along with the suggestion by Hambleton (1978) that several groups of judges representing different perspectives be included in making standard-setting decisions.

While recognizing that all standard-setting is judgmental and, by definition, arbitrary, a number of researchers feel that Glass's concerns (1978) are exaggerated. For example, Scriven (1978), while admitting the arbitrariness and imprecision of scoring standards, suggests that rather than no standards, we should continue our efforts to improve the process. He proposes that we acknowledge a substantial gray area around our test scores and that we categorize test performance in terms of probable mastery, uncertainty, and probable lack of mastery.

Linn (1978) also feels that explicit standards, despite their tentativeness, are preferable to nothing because "nothing really means hidden or unknown standards" (p. 307); while Block (1978) notes that some of the arbitrariness in test standards is "educationally healthy" and provides the opportunity to involve the public in decisions formerly left strictly to school personnel.

A variety of standard-setting methods is in existence today. An early review of these methods was conducted by Millman (1973). This was followed

up in 1976 by Meskauskas, who reviewed the models for setting pass-fail points in terms of their underlying concepts of mastery. Models were divided into two broad categories: those that viewed mastery as an area on a continuum (continuum models) and those that viewed mastery on an all-or-none basis (state models). Most recently, Hambleton and Eignor (1980) reviewed 19 procedures for setting competency test standards and recommended several particularly promising methods for further consideration. They acknowledged the fact, however, that at present there are few procedural guides for applying these standard-setting models.

Apart from the many technical/mathematical considerations that should be taken into account in selecting standard-setting methods, a crucial human (and legal) concern must be the consequence of the inevitable errors in test cut-off scores. We could pass a student who really doesn't deserve to pass, i.e., who doesn't have the required mastery of important skills (a "false-positive" error); or we could fail the student who really possesses the necessary skills (a "false-negative" error). As Livingston (no date) states, "The best we can do is to minimize the total harm from our errors" (p. 87). Shepard (1979, p. 66) notes that the seriousness of these errors will vary with the situation.

> When individuals are certified to practice in various professions as doctors, lawyers, teachers, the cost to society is much greater for the false-positives than for the false negatives. In these cases, relatively stringent standards ought to be set to protect the public against unqualified practitioners. The cost to individuals who are thereby unfairly failed is outweighed by the public good. In many instructional settings, however, the reverse is true.

Shaycoft has pointed out an interesting (and rather surprising) paradox in setting cutting scores on criterion-referenced tests (1979). She states that the cutting score on a particular test should not necessarily be the same as the standard of competence for an entire domain of knowledge of which the test is a sample. She asserts further--and provides the mathematics to support her assertion--that

> the more competent the group as a whole, the lower the cutting score may be set. Thus, if a pretest is being given to determine which examinees have achieved a sufficiently high level of competence to be exempted from a particular course,

> the cutting score should be set higher than it is for the posttest upon completion of the course. This is the finding when the aim is to find the cutting score that will minimize misclassifications (false passes and false failures). (pp. 70-71)

Technical issues are not the only problems that must be faced in setting performance standards. Schools are more and more facing the threat of legal action related to these standards, as discussed next.

Legal Considerations in Setting Performance Standards

As Spirer (no date) notes, "The institutionalization of performance testing in vocational education programs, for example, brings with it a series of legal concerns to which teachers and administrators must be sensitive" (p. 89). Tractenberg (no date), in his overview of the legal implications of performance testing in vocational education, stresses that these legal concerns "should play a significant role in the development of performance testing" (p. 96). To organize his presentation, he cites Brickell's Keynotes of Competency Testing (1978) as adapted to vocational education by Ahmann (1979). Of the eight keynotes that he lists, four relate explicitly to setting test-score standards:

- the number of proficiency standards that will be set
- the level(s) at which these standards will be set
- whether the standards will be for school programs or students
- the consequences for failing to achieve the standards

Two recommended courses of action come through rather clearly in Tractenberg's paper. In terms of level at which proficiency standards are set, he recommends "as a practical matter, unless a particular program is specifically designed to equip its students for journeyman positions, the standards should be geared to entry-level positions. The more important issue is likely to be whether the standards actually relate to the marketplace" (p. 101). In his discussion of the consequences of failing to achieve the test standards, Tractenberg recommends that

> The preferable, and in some cases the required, response to evidence that particular students had failed to meet proficiency is to direct appropriate educational assistance to

them. This may take the form of remediation for the individual students; it may involve broader programmatic or personnel responses. Surely, if a substantial percentage of the school's or program's students is failing to meet statewide or local standards, the overall educational program, including the quality of instructional staff, should be evaluated and perhaps upgraded. (p. 102)

Pullin (no date, p. 118) raises yet another legal issue relative to test scores. In her consideration of privacy and confidentiality in performance testing, she recommends that

- Test scores should not be disclosed to persons outside the school or to those not directly involved with the student's training without consent.
- Test scores should not be divulged to potential employers without the written consent of the parent or, if the student is over 18, the student.
- Interpretation of test results should be made available to students' parents.
- Tests should not include questions that unnecessarily infringe on students' privacy.

Tractenberg (no date, p. 103), looking toward future developments in the area of legal issues surrounding performance testing in vocational education, makes the following important recommendation:

> Vocational educators should not simply sit back and wait to be sued. They should deal in some preventive maintenance--they should attempt to head off legal challenges by fashioning and implementing performance testing programs in the most careful manner possible. If they do so, the law and the courts will have been an important partner in educational and professional reform.

Epilogue

Because so many of the readers of this report may find that the legal profession is often at their side (or perhaps "on their backs") as they implement competency testing in the schools, Tractenberg's advice merits special attention. While this report was not directed primarily at preventing law suits against test administrators, we hope that it provides the kind of "preventive maintenance" information that will be useful for improving the quality of competency testing in the vocational area.

REFERENCES

Abramson, T. Issues and models in vocational education evaluation. In T. Abramson, C. K. Tittle, & L. Cohen (Eds.), Handbook of vocational education evaluation. Beverly Hills, Calif.: Sage, 1979.

Abramson, T., & Vogrin, A. Instructional support systems. In T. Abramson, C. K. Tittle, & L. Cohen (Eds.), Handbook of vocational education evaluation. Beverly Hills, Calif.: Sage, 1979.

Abramson, T., Tittle, C. K., & Cohen, L. (Eds.). Handbook of vocational education evaluation. Beverly Hills, Calif.: Sage, 1979.

Agee, P. R. Florida moves to competency testing. School Shop, 1980, 39(8), 62.

Ahmann, J. S. Implications of the minimum competency testing movement for performance testing in vocational education. An unpublished paper, 1979.

American Association of Community and Junior Colleges and the American Vocational Association. Cooperation in vocational education: Executive summary. Arlington, Vir.: Authors, 1978.

American Institutes for Research. The definition of clinical competence in medicine: Performance dimensions and rationales for clinical skill areas. Palo Alto, Calif.: Author, 1976.

Ammerman, H. L. Performance content for job training: Stating the tasks of the job (Vol. 2). Columbus, Ohio: Ohio State University, 1977. (a)

Ammerman, H. L. Performance content for job training: Identifying relevant job performance (Vol. 3). Columbus, Ohio: Ohio State University, 1977. (b)

Ammerman, H. L., & Essex, D. W. Performance content for job training: Deriving performance requirements for training (Vol. 4). Columbus, Ohio: Ohio State University, 1977.

Ammerman, H. L., & Pratzner, F. C. Performance content for job training. Volume 1: Introduction. Research and Development Series No. 121. Columbus, Ohio: Ohio State University, Center for Vocational Education, March 1977. (ED 146 369)

Anderson, A. V., Laabs, G. J., Pickering, E. J., & Winchell, J. D. A personnel readiness training program: Final report. San Diego, Calif.: Navy Personnel Research and Development Center, 1977. (NPRDC TR 77-39)

Beach, D. P. Affective work competencies, Phase II. Final report. Columbia: University of Missouri, College of Education, 1978. (ED 167 727)

Bell, J. D., & Pickering, E. J. Use of performance measurement data from the 14A2 ASW team trainer complex in a performance proficiency assessment system. San Diego, Calif.: Navy Personnel Research and Development Center, 1979. (NPRDC TR 79-15)

Benson, A. Mid-America outreach. American Vocational Journal, 1978, 53(5), 26-28.

Blank, W. E. Competency-based education: Is it really better? School Shop, 1980, 39(8), 31-34.

Block, J. H. (Ed.). Mastery learning: Theory and practice. New York: Holt, Rinehart, & Winston, Inc., 1971.

Block, J. H. The "C" in CBE. Educational Researcher, 1978, 7(5), 13-16.

Bloom, B. S. Learning for mastery, UCLA Center for the Study of Evaluation. Evaluation Comment, 1968, 1(2).

Boyd, J. L., & Shimberg, B. Handbook of performance testing: A practical guide for test makers. Princeton, N. J.: Educational Testing Service, 1971.

Brickell, H. M. Seven key notes on minimum competency testing. Phi Delta Kappan, 1978, 59, 592-598.

Bunda, M. A., & Sanders, J. R. (Eds.). Practices and problems in competency-based education. Washington, D.C.: National Council on Measurement in Education, 1979.

Buros, O. K. The Eighth Mental Measurement Yearbook, Vol. II. Highland Park, N.J.: The Gryphon Press, 1978.

Burton, N. W. Societal standards. Journal of Educational Measurement, 1978, 15(4), 263-271.

Canada Employment and Immigration. Generic skills. Ottawa, Canada: Author, 1979.

Carbine, M. Evaluations of vocational education. In A. Lecht (Ed.), Evaluating vocational education policies and plans for the 1970s. New York: Praeger, 1974.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Chalupsky, A. B. VoComp to develop CBVE measures. School Shop, 1980, 39(8), 66-67.

Chapman, J. C. Trade tests: The scientific measurement of trade proficiency. New York: Henry Holt and Company, 1921.

Christal, R. E. The United States Air Force Occupational Research Project. Brooks Air Force Base, Tex.: Air Force Human Resources Laboratory, Air Force Systems Command, January 1974. (AD774574, NTIS)

Clearinghouse for Applied Performance Testing. Alternative conceptions of competence assessment: Proceedings of the 1979 Annual CAPT Conference. Portland, Ore.: Northwest Regional Educational Laboratory, November 1979.

Cockrell, J. T., & Kristiansen, D. M. Television as stimulus input in synthetic performance testing. Technical Paper 303. Alexandria, Vir.: U.S. Army Research Institute for the Behavioral and Social Sciences, August, 1978.

Conaway, L. E. Setting standards in competency-based education: Some current practices and concerns. In M. A. Bunda & J. R. Sanders (Eds.), Practices and problems in competency-based education. Washington, D.C.: National Council on Measurement in Education, 1979.

Cross, A. A. (Ed.). Vocational instruction. Arlington, Vir.: The American Vocational Association, Inc., 1980.

Darcy, R. L. Vocational education outcomes: Perspective for evaluation. Columbus, Ohio: The Ohio State University, The National Center for Research in Vocational Education, 1979.

Datta, L. Better luck this time. In T. Abramson, C. K. Tittle, & L. Cohen (Eds.), Handbook of vocational education evaluation. Beverly Hills, Calif.: Sage, 1979.

Denton, W. T. Program evaluation in vocational and technical education (Information Series No. 98). Columbus, Ohio: ERIC Clearinghouse on Vocation and Technical Education, 1973.

Dunn, J. Alternative systems of evaluation. Paper presented at the sixth annual seminar of New York State Directors of Occupational Education, 1978.

Dunnette, M. D. (Ed.). Handbook of industrial psychology. Chicago: Rand McNally College Publishing Company, 1976.

Erickson, R. C. Techniques for appraising vocational education. In A. A. Cross (Ed.), Vocational instruction: 1980 Yearbook. Washington, D.C.: American Vocational Association, 1979.

Erickson, R. C., & Wentling, T. L. Measuring student growth: Techniques and procedures for occupational education. Boston: Allyn and Bacon, 1976.

Fink, C. D., Wagner, H., Behringer, R. D., & Hibbits, F. L. A state-of-the-art review of techniques and procedures for the measurement of complex human performance. Consulting report. Alexandria, Vir.: Human Resources Research Organization, April 1974. (ED 150 167)

Finkle, R. B. Managerial assessment centers. In M. D. Dunnette (Ed.), Handbook of industrial psychology. Chicago: Rand McNally College Publishing Company, 1976.

Fitzpatrick, R., & Morrison, E. G. Performance and product evaluation. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Flanagan, J. C. Contributions of research in the armed forces to personnel psychology. Personnel Psychology, 1948, 1, 53-62.

Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

Foley, J. P., Jr. Criterion referenced measures of technical proficiency in maintenance activities. Wright-Patterson Air Force Base, Ohio: Air Force Human Resources Laboratory, October 1975. (AFHRL-TR-61)

Ford, J. P., Campbell, R. C., & Harris, J. H. Development of guidelines for administration of Skill Qualification Tests. Alexandria, Vir.: Human Resources Research Organization, Western Div., Fort Knox Office under Project ADTEST, July 1976. (HumRRo FR-WD-KY-76-9)

Gagné, R.(Ed.), Psychological principles in system development. New York: Holt, Rinehart and Winston, Inc., 1962.

Gampert, R. D. The assessment of clinical problem solving using a written simulation test. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Glaser, R., & Klaus, D. J. Proficiency measurement: Assessing human performance. In R. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart and Winston, Inc., 1962.

Glass, G. V. Standards and criteria. Journal of Educational Measurement, 1978, 15(4), 237-261.

Grant, G. Implications of competence-based education. In G. Grant et al. (Eds.), On competence: A critical analysis of competence-based reforms in higher education. San Francisco: Jossey-Bass Publishers, 1979.

Grant, G., Elbow, P., Ewens, T., Gamson, Z., Kohli, W., Neumann, W., Olesen, V., & Riesman, D. (Eds.). On competence: A critical analysis of competence-based reforms in higher education. San Francisco: Jossey-Bass Publishers, 1979.

Haller, E. J. Bringing the people back in: Emergencies, mistakes and boredom. VocEd, October 1978, 31-35.

Hambleton, R. K. On the use of cut-off scores with criterion-referenced tests in instructional settings, Journal of Educational Measurement, 1978, 15(4), 277-290.

Hambleton, R. K., & Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14(2), 75-96.

Hambleton, R. K., & Eignor, D. R. Competency test development, validation, and standard setting. In R. M. Jaeger & C. K. Tittle (Eds.), Minimum competency achievement testing: Motives, models, measures, and consequences. Berkeley, Calif.: McCutchan Publishing Corp., 1980.

Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 1978, 48(1), 1-47.

Hirst, B. A., Jr. The components of competency-based vocational education. American Vocational Journal, 1977, 52(8), 32-35.

Hogan, D. B. Is licensing public protection or professional protectionism? In P. S. Pottinger & J. Goldsmith (Eds.), Defining and measuring competence. San Francisco: Jossey-Bass Inc., 1979.

Hoyman, R. D. The development of a series of recommendations and procedures required in administration of a centralized program of occuoational competency assessment. University Park: Pennsylvania State University, Div. of Occupational & Vocational Studies, May 1978. (ED 160 754)

Huntington Beach Union High School District. Competency based model for vocational education programs. Huntington Beach, Calif.: Author, 1976.

Ingram, M. CBVE work at El Paso. School Shop, 1980, 39(8), 46-47.

Jaeger, R. M. Measurement consequences of selected standard-setting models. In M. A. Bunda & J. R. Sanders (Eds.), Practices and problems in competency-based education. Washington, D.C.: National Council on Measurement in Education, 1979.

Kazanas, H. C. Affective work competencies for vocational education. Information Series No. 138. Columbus, Ohio: ERIC Clearinghouse on Adult, Career, and Vocational Education, 1978. (ED 166 420)

Kelly, J. F., & Law, G. F. The national network: An assessment after five years. American Vocational Journal, 1978, 53(5), 35-37.

Klein, R. Some selected technical issues related to performance testing. In J. E. Spirer (Ed.), Performance testing: Issues facing vocational education. Research and Development Series No. 190. Columbus, Ohio: The Ohio State University, National Center for Research in Vocational Education, n.d. (ED 187930)

Klemp, G. O., Jr. Identifying, measuring, and integrating competence. In P. S. Pottinger & J. Goldsmith (Eds.), Defining and measuring competence. San Francisco: Jossey-Bass Inc., 1979.

Knaak, W. C. The anatomy of a competency-based school. American Vocational Journal, 1977, 52(8), 36-40.

Knaak, W. C. The significance of competency training. School Shop, 1980, 39(8), 48-53.

Knapp, J., & Sharon, A. A compendium of assessment techniques. Princeton, N.J.: Cooperative Assessment of Experiential Learning, Educational Testing Service, January 1975.

Laabs, G. J., & Kissler, G. D. Reliability and validity of critical task identification using a card-sort technique. Technical Note 78-16. San Diego, Calif.: Navy Personnel Research and Development Center, August 1978. (NPRDC TN 78-16)

Lecht, A. (Ed.). Evaluating vocational education policies and plans for the 1970s. New York: Praeger, 1974.

Lee, A. M. Project Baseline: Historiographic foundations for vocational education statistics. Education Researcher, 1977, 67, 3-9.

Lindquist, E. G. Educational measurement. Washington, D.C.: American Council on Education, 1951.

Linn, R. L. Demands, cautions, and suggestions for setting standards. Journal of Educational Measurement, 1978, 15(4), 301-308.

Livingston, S. A. Comments on the technical issues in performance testing. In J. E. Spirer (Ed.), Performance testing: Issues facing vocational education. Columbus, Ohio: The Ohio State University, National Center for Research in Vocational Education, n.d.

Lord, F. M. A theory of test scores. Psychometric Monograph No. 7. Psychometric Society, 1952.

Lloyd, B. H., & Hoover, H. D. Vertical equating using the Rasch model. Journal of Educational Measurement, 1980, 17(3), 179-193.

Maier, M. H., & Hirshfeld, S. F. Criterion-referenced job proficiency testing: A large scale application. Research Report 1193. Alexandria, Vir.: U.S. Army Research Institute for the Behavioral and Social Sciences, February 1978. (ADA055664, NTIS)

Meskauskas, J. A. Evaluation models for criterion-reference testing: Views regarding mastery and standard setting. Review of Educational Research, 1976, 46(1), 133-158.

McClelland, D. C. Testing for competence rather than for "intelligence." American Psychologist, 1973, 28(1), 1-14.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology Monograph, 1972, 56(4), 347-368.

McGuire, C. Simulation as an evaluation technique. In The Clearinghouse for Applied Performance Testing, Alternative conceptions of competence assessment: Proceedings of the 1979 Annual CAPT Conference. Portland, Ore.: Northwest Regional Educational Laboratory, November 1979.

McGuire, C. H., Solomon, L. M, & Bashook, P. G. Construction and use of written simulations. New York: The Psychological Corporation, 1976.

McKinney, F. Program evaluation in vocational education: A review. Information Series No. 117. Columbus, Ohio: The Center for Vocational Education, 1977.

McKinney, F. L., Gray, K. E., & Abram, M. (Eds.). Interpreting outcome measures in vocational education: A final report. Columbus, Ohio: The Ohio State University, The National Center for Research in Vocational Education, 1979.

Millman, J. Passing scores and test lengths for domain-referenced measures Review of Educational Research, 1973, 43(2), 205-216.

Morrell, J. Followup research as an evaluation strategy: Theory and methodology. In T. Abramson, C. K. Tittle, & L. Cohen (Eds.), Handbook of vocational education evaluation. Beverly Hills, Calif.: Sage, 1979.

Nelson, R. E. Survival skills: Mastering the human aspects of work. American Vocational Journal, 1977, 52(8), 64-66.

Nieva, V. F., Myers, D., & Glickman, A. S. An exploratory investigation of the Skill Qualification Testing system. Technical Report 390. Washington, D.C.: Response Analysis Corporation, Advanced Research Resources Organization, 1979.

Oen, U. T. Step by step to a CBVE program. School Shop, 1980, 39(8), 35-37.

Office of Strategic Services, Assessment Staff. Assessment of men. New York: Holt, Rinehart and Winston, 1948.

Ohio Department of Education. The Ohio Vocational Achievement Test Program. Columbus, Ohio: The Ohio State University, Instructional Materials Laboratory, 1979.

Oliver, J. D. Measuring student competencies--Heart of CBVE. American Vocational Journal, 1978, 53(1), 48-53.

Olivo, C. T. NOCTI: Precursors and plans. School Shop, 1980, 39(8).

Olivo, C. T., Panitz, A., Schaefer, C., Nelson, R., & Barlow, M. A consortium for occupational competency testing of trade and industrial/technical teachers. Final report. New Brunswick, N.J.: Rutgers State University, Graduate School of Education, June 1973. (Grant Number OEG2-9-100474-1044)

Olson, P. A., & Freeman, L. Defining competency in teacher licensing usage. In P. S. Pottinger & J. Goldsmith (Eds.), Defining and measuring competence. San Francisco: Jossey-Bass Inc., 1979.

Orlich, D. C., Andersen, D. G., Dodd, C. C., Baldwin, L., & Ohrt, B. A. Evaluation models for vocational educators. Olympia, Wash.: Washington State Commission for Vocational Education, 1978.

Osborn, W. C., Campbell, R. C., Ford, J. P., Hirshfeld, S. F., & Maier, M. H. Handbook for the development of skill qualification tests. Alexandria, Vir.: Human Resources Research Organization, November 1977. (Report No. P-77-5)

Owens, T. R. Experience based learning and assessment in public school systems. In The Clearinghouse for Applied Performance Testing, Alternative conceptions of competence assessment: Proceedings of the 1979 Annual CAPT Conference. Portland, Ore.: Northwest Regional Educational Laboratory, November 1979.

Panitz, A., & Olivo, C. T. National Occupational Competency Testing Project: Handbook for developing and administering occupational competency tests. Volume 3. New Brunswick, N.J.: Rutgers, the State University, Department of Vocational-Technical Education, February 1971. (ED 057 217)

Pickering, E. J., & Anderson, A. V. Measurement of job-performance capabilities. San Diego, Calif.: Navy Personnel Research and Development Center, December 1976. (ADA033992, NTIS)

Pope, L. T., & Meister, D. (Eds.) Symposium proceedings: Productivity enhancement: Personnel performance assessment in Navy systems. San Diego, Calif.: Navy Personnel Research and Development Center, 1977.

Popham, W. J. Educational evaluation. Englewood Cliffs, NJ: Prentice-Hall, 1975.

Popham, W. J. As always, provocation. Journal of Educational Measurement, 1978, 15(4), 297-300.

Pottinger, P. S., & Goldsmith, J. (Eds.). Defining and measuring competence. San Francisco: Jossey-Bass Inc., 1979.

Pratzner, F. C. Job analysis: Key to program articulation. Educational Technology, March 1979, 30-34.

Pullin, D. Performance testing in vocational education--Lessons to be learned from the minimum competency testing movement. In J. E. Spirer (Ed.), Performance testing: Issues facing vocational education. Columbus, Ohio: The Ohio State University, National Center for Vocational Education, n.d.

Riesman, D. Society's demands for competence. In G. Grant et al. (Eds.), On competence: A critical analysis of competence-based programs in higher education. San Francisco: Jossey-Bass Publishers, 1979.

Russell, E. B. Examples of CBVE and where to find them. American Vocational Journal, 1978, 53(5), 55-58.

Sanders, J. R., & Sachse, T. P. Applied performance testing in the classroom. Journal of Research and Development in Education, 1977, 10(3), 92-104.

Scriven, M. How to anchor standards. Journal of Educational Measurement, 1978, 15(4), 273-275.

Sharon, A. T. Assessing occupational competences--A CAEL handbook. Princeton, N.J.: Educational Testing Service, 1977.

Shaycoft, M. F. Handbook of criterion-referenced testing: Development, evaulation, and use. New York: Garland STPM Press, 1979.

Shepard, L. A. Setting standards. In M. A. Bunda & J. R. Sanders (Eds.), Practices and problems in competency-based education. Washington, D.C.: National Council on Measurement in Education, 1979.

Shriver, E. L., & Foley, J. P., Jr. Evaluating maintenance performance: The development and tryout of criterion referenced iob task performance tests for electronic maintenance. Wright-Patterson Air Force Base, Ohio: Air Force Human Resources Laboratory, September 1974. (AFHRL-TR-74-57(II), Pt. 1)

Slater, S. J. Applied performance testing: Typology, advantages, limitations and examples. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, April 1980.

Slinde, J. A., & Linn, R. L. A note on vertical equating via the Rasch model for groups of quite different ability and tests of quite different difficulty. Journal of Educational Measurement, 1979, 16(3), 159-165.

Spirer, J. E. (Ed.) Performance testing: Issues facing vocational education. Research and Development Series No. 190. Columbus, Ohio: The Ohio State University, National Center for Research in Vocational Education, n.d.

Stromsdorfer, E. W. Review and synthesis of cost-effectiveness: Studies of vocational and technical education. Information Series No. 57. Columbus, Ohio: The Center for Vocational and Technical Education, 1972.

Taylor, C. M., Darcy, R. L., & Bolland, K. A. Vocational education outcomes: Annotated bibliography of related literature. Bibliography Series No. 48. Columbus, Ohio: The Ohio State University, The National Center for Research in Vocational Education, 1979.

Taylor, E. N., & Vineberg, R. Annex to "Performance Test Development for Skill Qualification Testing": Army performance tests - A critique (Final report). Alexandria, Vir.: Human Resources Research Organization, 1975.

Taylor, J. E., & Staff, ATC-PERFORM. Establishing the concepts and techniques of performance-oriented training in Army Training Centers: A summary report. Alexandria, Vir.: Human Resources Research Organization, Work Unit ATC-PERFORM, June 1975. (HumRRO-TR-75-21)

Thorndike, R. L. (Ed.). Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Tractenberg, P. L. Legal implications of performance testing in vocational education: An overview. In J. E. Spirer (Ed.), Performance testing: Issues facing vocational education. Research and Development Series No. 190. Columbus, Ohio: The Ohio State University, National Center for Research in Vocational Education, n.d.

Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14(2), 181-196.

U. S. Department of Health, Education, and Welfare, Office of Education. Annual evaluation report on programs administered by the U.S. Office of Education, Fiscal Year 1979. Prepared by the Office of Evaluation and Dissemination, U. S. Department of Health, Education, and Welfare.

Venn, G. Criteria against which vocational education should be held accountable. In F. L. McKinney, K. E. Gray, & M. Abram (Eds.), Interpreting outcome measures in vocational education: A final report. Columbus, Ohio: The Ohio State University, The National Center for Research in Vocational Education, 1979.

Waag, W. L., & Knoop, P. A. Planning for aircrew performance measurement R&D: U.S. Air Force. In L. T. Pope & D. Meister (Eds.), Symposium proceedings: Productivity enhancement: Personnel performance assessment in Navy systems. San Diego, Calif.: Navy Personnel Research and Development Center, 1977.

Walejko, C. Managing CBVE in the classroom. American Vocational Journal, 1977, 52(9), 34-36.

Wentling, T. L. The assessment of learning: A base for instructional evaluation. In A. A. Cross (Ed.), Vocational instruction. Arlington, Vir.: The American Vocational Association, Inc., 1980.

Wentling, T. L., & Lawson, T. E. Evaluating occupational education and training programs. Boston, Mass.: Allyn and Bacon, 1975.

Willingham, W. W. Principles of good practice in assessing experiential learning. Princeton, N. J.: Educational Testing Service, Cooperative Assessment of Experiential Learning, 1977.

Wright, B. D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14(2), 97-116.